Measuring precise radial velocities and cross-correlation function line-profile variations using a Skew Normal density

U. Simola X. Dumusque Jessi Cisewski-Kehe

May 30, 2018

Abstract

Stellar activity is one of the primary limitations to the detection of low-mass exoplanets using the radial-velocity (RV) technique. Stellar activity can be probed by measuring time dependant variations in the shape of the cross-correlation function (CCF), often estimated using the different moments of the modelled CCF. Therefore estimating the moments of the CCF with high precision is essential to de-correlate exoplanet signals from spurious RV signals originating from stellar activity. We propose to estimate the moments of the CCF by fitting a model using a Skew Normal (SN) density shape, which unlike the commonly employed Normal density, [[**Xavier:** includes an skewness parameter to capture the asymmetry of the CCF induced by stellar activity, but also the natural asymmetry induced by convective blueshift.]] The performance of the proposed method is compared to the Normal density using both simulated and real observations with varying levels of activity and signal-to-noise ratio (SNR) levels. When considering the real observations, the correlation between the RV's and the asymmetry of the CCF and the correlation between the RV's and the width of the CCF are stronger when using the parameters derived from the SN than the Normal approach. This suggests that the CCF asymmetry and the CCF width derived using a SN may be more sensitive to stellar activity, which can be helpful with estimating stellar rotational periods and generally characterizing the stellar activity signals. The estimated uncertainties in the estimated RV's using the proposed SN approach are on average 10% smaller than the uncertainties calculated on the mean of the Normal, and the estimated uncertainties on the SN asymmetry parameter are on average 15% smaller than the commonly used Bisector Inverse Slope Span (BIS SPAN) approach for estimating the asymmetry of the CCF. The derived benefits of using the proposed SN approach includes an asymmetry parameter along with several options for estimating the RV, while the Normal model only includes an RV estimate and requires a separate procedure for estimating the asymmetry. Furthermore, the estimated uncertainties on the RV's and the asymmetry parameter are smaller in the SN setting.

1 Introduction

When working with radial velocities (RV's), one of the main limitations to the detection of small-mass exoplanets is no longer the precision of the instruments used, but the different sources of variability induced by the stars (e.g. Feng et al. 2017; Dumusque et al. 2017; Rajpaul et al. 2015; Robertson et al. 2014). Stellar oscillations, granulation phenomena, and stellar activity can all induce apparent RV signals (e.g. Saar and Donahue 1997; Queloz et al. 2001; Desort et al. 2007; Dumusque et al. 2011; Dumusque 2016) that are above the meter-per-second precision reached by the best high-resolution spectrographs (HARPS, HARPS-N, Mayor et al. 2003; Cosentino et al. 2012). It is therefore mandatory to better understand stellar signals and to develop methods to correct for them, if in the near future we want to detect or confirm an Earth-twin planet using the RV technique. This is even more true now that instrument like the Echelle SPectrograph for Rocky Exoplanet and Stable Spectroscopic Observations (ESPRESSO) (Pepe et al. 2014) and EXtreme PREcision Spectrometer (EXPRES) (Fischer et al. 2016) should have the stability to detect such signals. However, if solutions are not found to mitigate the impact of stellar activity, the detection or confirmation of potential Earth-twins will be extremely challenging and false detections could plague the field.

One of the most challenging stellar signal to characterize and to correct for is the signal induced by stellar activity. Stellar activity is responsible for creating magnetic regions on the surface of stars, and those regions change locally the temperature and the convection, which can induce spurious RV's variations (Meunier et al. 2010; Dumusque et al. 2014). In theory, it should be easy to differentiate between the pure Doppler-shift induced by a planet, which shifts the entire stellar spectrum, and stellar activity, which modifies the shape of spectral lines and by doing so create a spurious shift of the stellar spectrum (Saar and Donahue 1997; Hatzes 2002; Kurster et al. 2003; Lindegren and Dravins 2003; Desort et al. 2007; Lagrange et al. 2010; Meunier et al. 2010; Dumusque et al. 2014). However, on quiet GKM dwarfs, the main target for precise RV's measurements, stellar activity can induce signals of a few $\mathrm{m\,s^{-1}}$. This corresponds physically to variations smaller than 1/100th of a pixel on the detector making the changing shape of the spectral lines challenging to detect.

In order to measure such tiny variations, a common approach is to average the information of all the lines in the spectrum by cross correlating the stellar spectrum with a synthetic (Baranne et al. 1996; Pepe et al. 2002) or an observed stellar template (Anglada-Escudé and Butler 2012). The result of this operation gives us the cross-correlation function (CCF). The CCF gives the spectrum's cross-correlation with the template as the template is shifted according to different RVs. To measure the Doppler-shift between different spectra and therefore to retrieve the RV's of a star as a function of time, the variations of the CCF barycenter are calculated. The barycenter is generally estimated by fitting a Normal density to the CCF and retaining its mean. Variations in line shape between different spectra, which indicate the presence of signals induced by stellar activity, are measured by analyzing the different moments of the CCF. Usually, the width of the CCF is estimated using the full-width half-maximum (FWHM) of the fitted Normal density, and its asymmetry using the the bisector inverse slope span (BIS SPAN, Queloz et al. 2001).

If an apparent RV signal is induced by activity, generally a strong correlation will be observed between the RV and chromospheric activity indicators like log(R'_{HK}) or H-α (Boisse et al. 2009; Dumusque et al. 2012; Robertson et al. 2014), but also between the RV and the FWHM of the CCF or its BIS SPAN (Queloz et al. 2001; Boisse et al. 2009; Queloz et al. 2009; Dumusque 2016). It is therefore common now, that when fitting a Keplerian signal to a set of RVs to look for a planet, the model includes in addition linear dependancies with the log(R'_{HK}), the FWHM and the BIS SPAN (Dumusque et al. 2017; Feng et al. 2017). It is also common to add a Gaussian process to the model to account for the correlated noise induced by stellar activity. The hyperparameters of the Gaussian process can be trained on different activity indicators (Haywood et al. 2014; Rajpaul et al. 2015). It is therefore essential for mitigating stellar activity to obtain activity indicators that are the most correlated with the RV's but also for which we can obtain the best precision.

Several indicators have been developed that are more sensitive to line asymmetry than the BIS SPAN. In Boisse et al. (2011), the authors develop V_{span}, which is the difference between the RV measured respectively by fitting a Normal density to the upper and the bottom part of the CCF. This CCF asymmetry parameter is shown to be more sensitive than the BIS SPAN at low signal-to-noise ratio (SNR). Figueira et al. (2013) studied the use of two new indicators, bi-Gauss and V_{asy}. The authors were able to show that when using bi-Gauss, the amplitude in asymmetry is 30% larger than when using BIS SPAN, therefore allowing the detection of lower levels of activity. They also demonstrated that V_{asy} seems to be a better indicator of line asymmetry at high SNR, as its correlation with RV is more significant than any of the previously proposed asymmetry indictors.

In all the methods described above, except bi-Gauss, the RV and the FWHM are derived using a Normal density fitted to the CCF, and the asymmetry is estimated using another approach. In this paper we propose to use a Skew Normal (SN) density to estimate with a single fit of the CCF, the RV, the FWHM and the asymmetry of the CCF, as this function includes a skewness parameter (Azzalini 1985).

The paper is organized as follow. In Sec. 2 we introduce the SN density, describe its applicability for modeling the CCF, and study how the SN parameters relate to the RV, FWHM and BIS SPAN of the CCF. In Sec. 3 we propose an expanded linear model used to correct the estimated RV's from the inferred stellar activity, which extends the linear models previously proposed for this purpose (Dumusque et al. 2017; Feng et al. 2017). In Sec. 4 the performance of the SN fit to the CCF is investigated using simulations coming from the Spot Oscillation And Planet (SOAP) 2.0 (Dumusque et al. 2014), followed by an analysis of real observations in Sec. 5. Sec. 6 considers derived error bars for the different estimated CCF parameters, and finally a discussion of the results and conclusions are included in Sec. 7 and Sec. 8, respectively.

2 The Skew Normal distribution

The Skew Normal (SN) distribution is a class of probability distributions which includes the Normal distribution as a special case (Azzalini 1985). The SN distribution has, in addition to a location and a scale parameter analogous to the Normal distribution's mean and standard deviation, a third parameter which describes the skewness (i.e. the asymmetry) of the distribution. Considering a random variable $Y \in \mathbb{R}$ (where $\mathbb{R}$ is the real line) which follows a SN distribution with location parameter $\xi \in \mathbb{R}$, scale parameter $\omega \in \mathbb{R}^+$ (i.e., the positive real line), and skewness parameter $\alpha \in \mathbb{R}$, its density at some value $y \in Y$ can be written as

$$SN(y; \xi, \omega, \alpha) = \frac{2}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\Phi\left(\frac{\alpha(y-\xi)}{\omega}\right), \tag{1}$$

where ϕ and Φ are respectively the density function and the distribution function of a standard Normal distribution[1]. The skewness parameter α quantifies the asymmetry of the SN. Examples of SN densities under different skewness parameter values and the same location and scale parameters ($\xi = 0$ and $\omega = 1$) are displayed in Fig. 1. A usual Normal distribution is the special case of the SN distribution when the skewness parameter α is equal to zero[2]. For reasons related to the

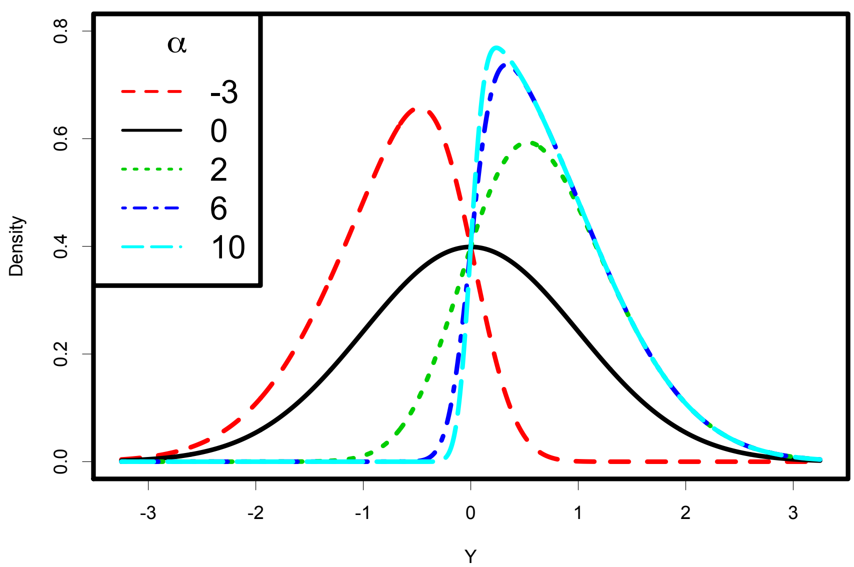

Figure 1: Density function of a random variable Y following the SN distribution $SN(\xi, \omega^2, \alpha)$ with location parameter $\xi = 0$, scale parameter $\omega = 1$ and different values of the skewness parameter α indicated by different colors and line types. Note that the solid black line has an $\alpha = 0$, making it a Normal distribution.

interpretation of the parameters in Eq. 1 and computational issues with estimating α near 0, a different parametrization is used in this work, which is referred to as the *centered parametrization*

[1] A standard Normal distribution is a Normal distribution with a mean of 0 and a standard deviation of 1.

[2] This can be seen from Eq. 1. If $\alpha = 0$ then $\Phi\left(\frac{\alpha(y-\xi)}{\omega}\right) = \Phi(0) = 0.5$ and therefore $SN(y; \xi, \omega, 0) = \frac{1}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)$ which is the density of a Normal distribution. Note that $\Phi(0) = 0.5$ because $\Phi(0)$ is the the probability that a standard Normal random variable is less than or equal than 0.

(CP). This CP is much closer to the parametrization of a Normal distribution, as it uses a mean parameter μ, a variance parameter σ^2 and a skewness parameter γ. In order to define the CP, we need to express the CP parameters (μ, σ^2, γ) as a function of (ξ, ω^2, α). This can be done using the following relations:

$$\mu = \xi + \omega\beta, \quad \sigma^2 = \omega^2(1-\beta^2), \quad \gamma = \frac{1}{2}(4-\pi)\beta^3\left(1-\beta^2\right)^{-3/2}, \tag{2}$$

where $\beta = \sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{\sqrt{1+\alpha^2}}\right)$ (e.g. Arellano and Azzalini 2010).

By using Eq. 2, the new set of parameters (μ, σ^2, γ) provides a clearer interpretation of the behavior of the SN distribution. For the α values used in Fig. 1, the corresponding values of (μ, σ^2, γ) are displayed in Table 1. In particular, μ and σ^2 are the actual mean and variance of the distribution, rather than simply a location and scale parameter, and γ provides an measure of the skewness of the SN. Along with the mean of the SN, we consider the median of the distribution as a measure of center, which is used in the proposed method. See Table 1 for the medians of the SN densities displayed in Fig. 1.

α	μ	σ^2	γ	Median
-3	-0.757	0.427	-0.667	-0.672
0	0.000	1.000	0.000	0.000
2	0.714	0.491	0.454	0.655
6	0.787	0.381	0.891	0.674
10	0.794	0.370	0.956	0.674

Table 1: CP values (μ, σ^2, γ) along with the median corresponding to the α values shown in Fig. 1, with location parameter $\xi = 0$ and scale parameter $\omega = 1$. Values are rounded to three decimal places.

Further details about the parametrization from Eq. 1, called the *Direct Parametrization* or DP, the CP, and general statistical properties of the SN are treated in rigorous mathematical and statistical viewpoints in the book by Azzalini and Capitanio (2014).

2.1 Fitting the Skew Normal density to the CCF

To fit the CCF using a SN density shape, we use a least-squares algorithm and the following model:

$$f_{CCF}(x_i) = \mathrm{C} - \mathrm{A} \times SN(x_i; \mu, \sigma^2, \gamma), \quad i = 1, \ldots, n \tag{3}$$

where C is an unknown offset for the continuum of the CCF, A is the unknown amplitude of the CCF, commonly referred to as the contrast, and μ, σ^2 and γ are the mean, variance and skewness of the SN as defined above. The values $x_1, \ldots, x_n$ are the different values of the x-axis of the CCF, generally in the unit of a velocity.

When fitting a Normal density to the CCF, the estimated mean of the model is used as the estimated RV, the FWHM of the Normal density[3] represents the width of the CCF. Because the Normal density is symmetric, the skewness is always equal to 0 so a separate approach is needed to estimate the skewness of the CCF. An estimated skewness parameter is generally obtained by calculating the BIS SPAN of the CCF (see Sect. 1, and e.g. Queloz et al. 2001).

With the proposed SN approach, we propose two estimators of the RV: the mean and median of the SN model fit (referred to as SN mean RV and SN median RV, respectively), and present advantages and limitations for both of these choices in Sec. 5 and Sec. 6. The width of the SN, SN FWHM, is defined in the same way as for the Normal density[4], and finally the skewness of the CCF is estimated by the γ parameter.

To evaluate the strength of the correlation between the estimated RV's and the different stellar activity indicators, we calculated the Pearson correlation coefficient, which in its general form is defined as:

$$R(x, y) = \frac{cov(x, y)}{\sigma(x), \sigma(y)}, \tag{4}$$

where x and y are two quantitative variables, $cov(x, y)$ indicates the covariance between x and y, and $\sigma(x)$ and $\sigma(y)$ represent their standard deviations. A $p-$value for the statistical test having null hypothesis $H_0 : R = 0$ is provided, along with a 95% confidence interval for R when needed.

3 Radial Velocity correction function for stellar activity

[[**Xavier:** Exoplanets only produce a pure RV signal. Stellar activity on the contrary, and in particular the presence of active regions on the stellar photosphere, does not produce a blueshift or redshift of the entire stellar spectra, but creates a spurious RV signal by modifying the shape of spectral lines. To track these variations in the shape of the spectral lines]], the general approach consists in using the FWHM, the BIS SPAN or the indicators introduced in Boisse et al. (2011) or Figueira et al. (2013), which provide information on the width and asymmetry of the CCF. A strong correlation between the estimated RVs and one or more of these parameters provides an indication that stellar activity signals are affecting our measurements.

[[**Xavier:** When fitting for planetary signals in RV data, it is common to include in the fitted model, linear dependencies with the BIS SPAN and the FWHM, to take into account the signal induced by stellar activity (e.g. Dumusque et al. 2017; Feng et al. 2017). In this paper, we propose to add additional parameters in the model to correct for stellar activity: first the amplitude parameter A of the CCF, generally referred to as the CCF contrast, and the interaction between the BIS SPAN and the FWHM (or γ and SN FWHM in the SN case). The stellar activity correction we propose can therefore be written as:]]

$$RV_{\text{stellar activity}} = \beta_0 + \beta_1 A + \beta_2 \gamma + \beta_3 \text{SN FWHM} + \beta_4 (\gamma \text{SN FWHM}) + \epsilon, \tag{5}$$

[3] FWHM $= 2\sqrt{2 \ln 2}\, \sigma$ with standard deviation σ

[4] Note that SN FWHM does not correspond to the width of the SN density at half maximum like in the Normal case.

where β_0 is the intercept and ϵ is the vector of the errors with mean equal to 0 and covariance matrix equal to $\sigma^2 I$ (I defined as the identity matrix). The inclusion of the contrast parameter A is justified by the fact that the presence of a spot on the stellar surface produces a changes in the amplitude of the CCF and not only on its asymmetry or width (see e.g. Fig. 2 in Dumusque et al. 2014). The reasons for including in Eq. 5 a variable that quantifies the interaction between γ and SN FWHM (or BIS SPAN and FWHM) will be better justified through the results of the examples presented in Sec.4. [[**Xavier:** This interaction term allows to take into account cases when for different values of SN FWHM (or FWHM), the linear coefficient β_2 for γ (or BIS SPAN) varies]]. [[**Xavier:** Umberto, Jessi, is this correct ?]]

In order to show the goodness of this correction, a statistical test on the parameters β_0, β_1, β_2, β_3 and β_4 is presented, where the null hypothesis is $H_0 : \beta_i = 0$, for $i = 0, \ldots, 4$. The level for not rejecting the null hypothesis is fixed to 0.05. The coefficient of multiple correlation R^2 is introduced in order to explain how well this linear combination addresses the variability of the RV variation induced by stellar activity.

When working with a linear regression, there are several ways to select the variables to include in the model. While usually the stepwise technique is used (Efroymson 1960; Hocking 1976), the proposed function defined in Eq. 5, that accounts for stellar activity, is the result of statistical and astronomical considerations. In particular we checked that the correlations between the proposed parameters were not approaching one: if it was the case, the matrix needed to calculate the estimates would be singular, hence non invertible. This problem is known in statistics with the term multicollinearity. A detailed discussion of the topic can be found in the book by Belsley (1991). [[**Xavier:** It is common to see some correlation between the amplitude parameter A and the FWHM (or SN FWHM) of the CCF. However, in the analysis of real data presented in this work, we never observed a correlation coefficient exceeding 0.66 and therefore, the problem of multicollinearity is avoided. Finally, we investigate as well the statistical significant of the interaction term between A and the width, and A and the asymmetry of the CC, however, those interaction were relevant for accounting for stellar signal.]]

4 Simulation Study

In order to evaluate the performance of the proposed SN approach for modelling the CCF and the benefit of using the proposed correction for stellar activity (See Eq. 5), we begin by considering a simulation study using spectra generated from the Spot Oscillation And Planet 2.0 code (SOAP 2.0, Dumusque et al. 2014).

[[**Xavier:** For a given configuration of spots and faculae on the stellar surface, SOAP 2.0 gives as output the simulated CCF as a function of rotational phase. The code returns as well the RV and the FWHM by fitting a Normal density to the CCF, and the BIS SPAN by calculating the bisector of the CCF. SOAP 2.0 gives us therefore noiseless CCFs affected by stellar activity, which will be used to compare the benefits of fitting a SN density to the CCF compared to a Normal density fit.]]

[[**Xavier:** For the simulations shown below, we modelled a star similar to the Sun. The stellar rotational period is set to 25.0 days, the radius to a solar radius and the star is seen equator on. The stellar effective temperature is set to 5778 K (NASA Planetary Fact Sheets), and we use a quadratic limb-darkening relation with linear and quadratic coefficients 0.29 and 0.34, respectively (Oshagh et al. 2013; Claret and Bloemen 2011). In addition, to be able to compare the result of those simulations with real data obtained with the HARPS spectrograph in Sect. 5, we asked SOAP 2.0 to generate CCFs of width 40 $\mathrm{km\,s^{-1}}$and of resolution 115'000, so that those CCFs would have the same properties as the one returned by the HARPS data reduction.]]

4.1 Faculae

[[**Xavier:** To see the impact of a faculae on the different parameters of the CCF, we simulated the effect of an equatorial faculae of size 3% relative to the visible stellar hemisphere. The faculae is face on when the phase equals to 0. Note that a 3% faculae is relatively big for the Sun. At maximum activity, big faculae have generally a size of 1%. In Fig. 2, we compare the barycentric variation of the CCF as measured when fitting a Normal density and taking its mean (RV), and when fitting a SN density and taking its mean (SN mean RV) or its median (SN median RV). We see that all the different estimates of the CCF barycenter present a signal of similar amplitude, however the signal obtained with SN mean RV is slightly different with respect to the two others, with a maximum amplitude happening at a different phase.]]

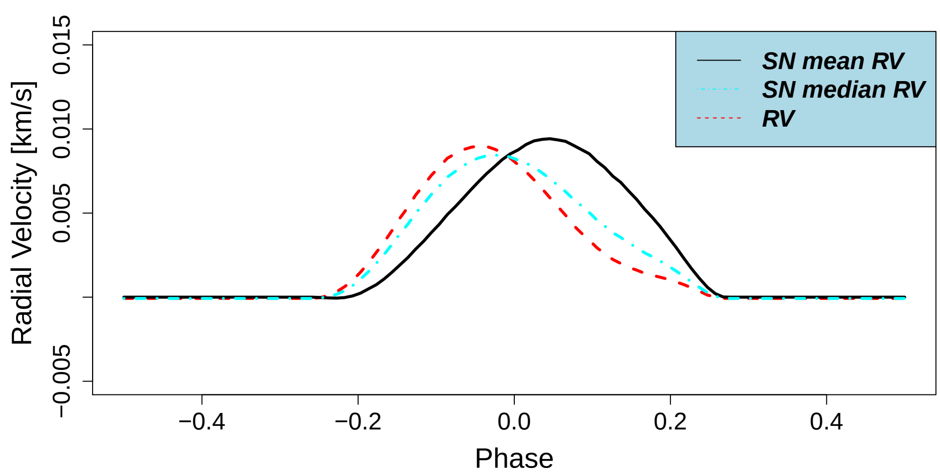

Figure 2: RV's changes as function of the orbital phase in the case in which a faculae is present on the photosphere of the star. SN mean RV seems to have the smallest spurious variations caused by the faculae.

[[**Xavier:** In Fig. 3, we compare the correlation between the different RV estimates and the different asymmetry or width estimates. As we can see, the strength of the correlation between

γ and SN mean RV, and γ and SN median RV are much stronger than the correlations between BIS SPAN and RV, with Pearson correlation coefficient R values of 0.46, -0.67 and -0.09, respectively. Regarding the width barycenter correlations, we also find a stronger correlation between SN FWHM and SN mean RV compared to the one between FWHM and RV, R = 0.98 and 0.84, respectively. In this case however, the correlation between SN FWHM and SN median RV is smaller with R = 0.50. This first analysis shows that in the case of a faculae, using some parameters from the SN can lead to much stronger correlation than the usual Normal parameters and therefore, the SN parameters seems to be better

probe better stellar activity. We will of course confirm this in the next sections presenting the case of a spot and of a spot plus a planet, but also in Sec 5 when we will apply the SN fitting to real data sets.]]

Since the RV variation displayed in Fig. 2 is caused by only stellar activity, in this case a faculae, [[**Xavier:** we applied the activity correction proposed in Eq. 5 to check its efficiency in the case of a faculae.]] The results of this correction are displayed in Fig. 4 and the statistical tests on the coefficients involved in Eq. 5 are summarized in Table 2. It is straightforward to see that the proposed correction for [[**Xavier:** stellar activity is able to account for the majority of the activity signal created by a faculae, with a R^2 of our model larger than 0.95. In addition, the rms of the different estimates of the RV reduces from about 3 $\mathrm{m\,s^{-1}}$before correction to values below 0.15 $\mathrm{m\,s^{-1}}$after correction. We see a slightly smaller rms after correction for the SN parameters, however the different is not significant. When comparing the correction proposed in Eq. 5 with what is generally used, i.e. a linear combination of the asymmetry and width parameter, we see that the proposed correction is able to reduce the rms of the RV residuals by a factor of 2. Looking at the significance of the coefficients in table 2, we observe that the intercept β_0 is only significant at a level of 1% in the case of the Normal parameters, the coefficient β_3 that corresponds to the CCF width is only significant when using the SN parameters, and the coefficient β_4 that account for the interaction between the width and the asymmetry parameters is never significant.]]

4.2 Spot

[[**Xavier:** In this section, we simulate the effect on the CCF parameters of an equatorial spot of size 1% relative to the visible stellar hemisphere. The spot is face on when the phase equals to 0. Note that this is an extremely big spot for the Sun, as in general big spots are more in the regime of 0.1%. In Fig. 5, we shows the barycentric variation of the CCF induced by this simulated spot. Contrary to the case of the faculae seen before, for the spots, all the different estimates of the CCF barycenter have exactly the same shape in variation. The amplitude for SN mean RV is however slightly smaller.]]

Fig. 6 shows the correlations between the asymmetry parameters and the different estimates for the CCF barycenter: SN mean RV, SN median RV and RV. [[**Xavier:** The correlation between γ and SN median RV is the strongest with a R = 0.94, followed then by the correlation BIS SPAN -RV and γ-SN mean RV, with R = 0.86. Regarding the correlation between the width and the CCF barycenter, we note that the variation is seen as a circle in this parameter space and therefore no

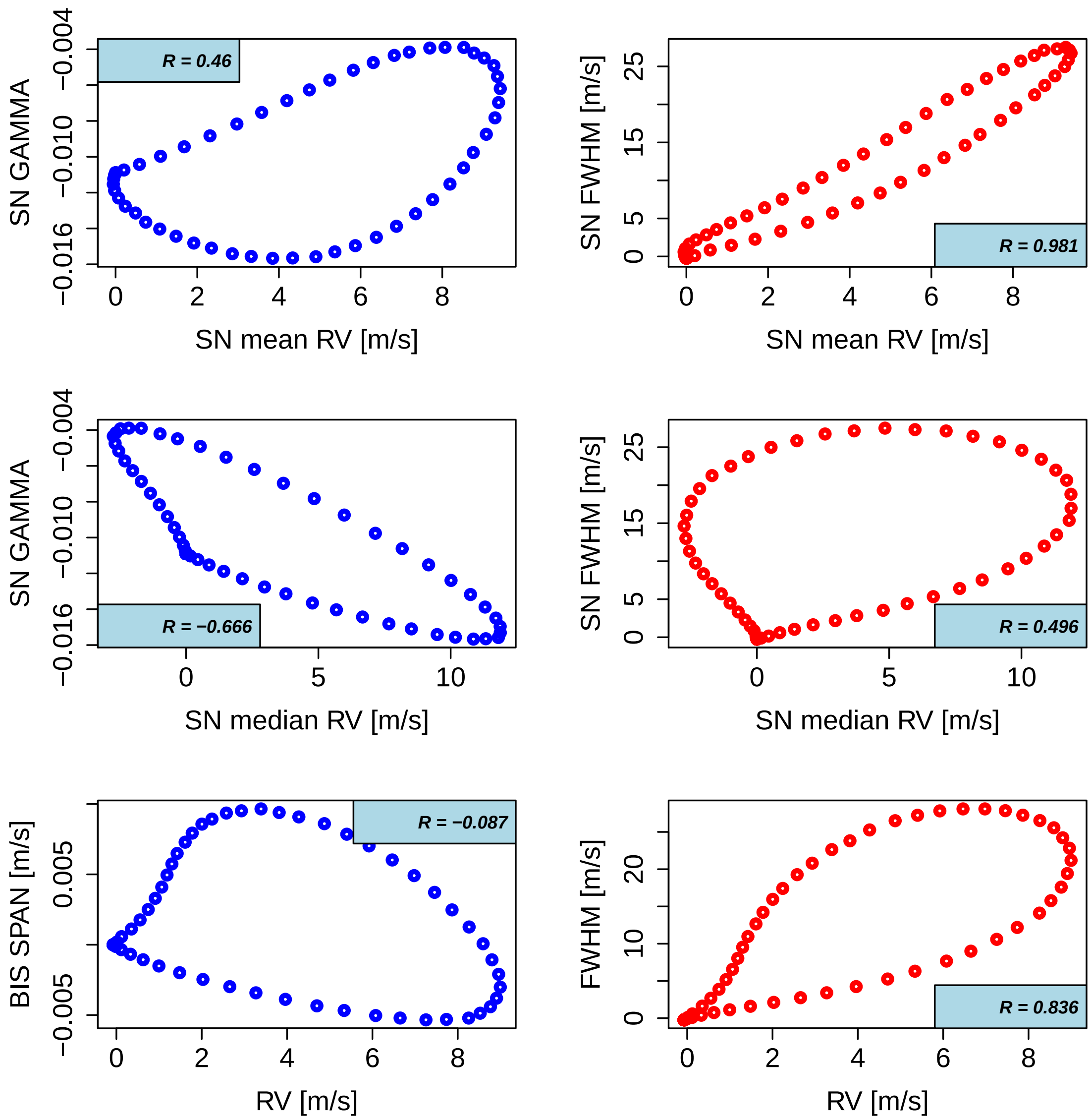

Figure 3: Evaluation of the correlation between the RV's and the asymmetry parameters when a faculae is present on the photosphere of the star. In this case both the shape and the width of the CCF changes as the faculae moves, producing statistically significative correlations between the RV's and respectively the asymmetry parameter and the width parameter.

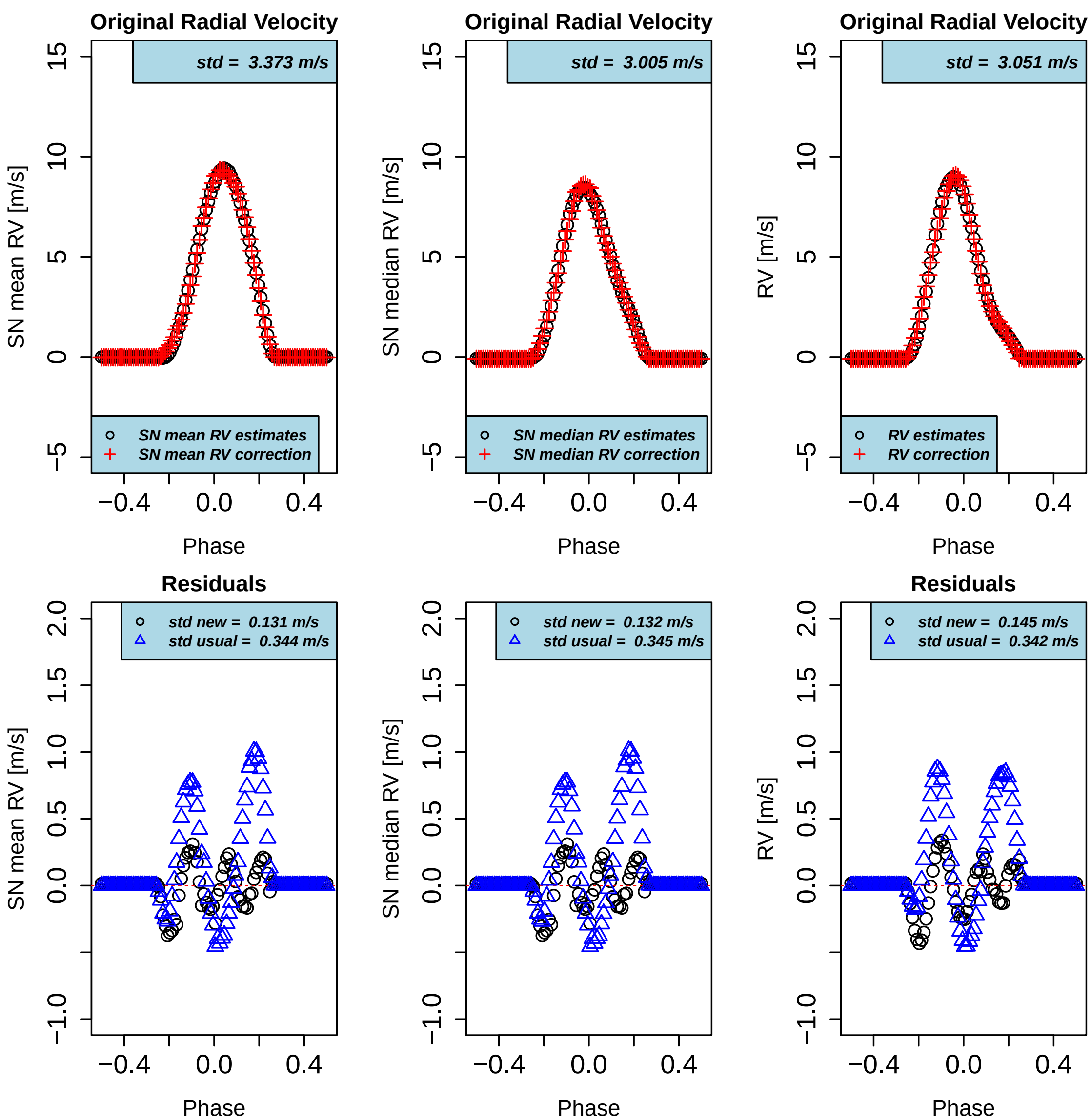

Figure 4: Set of spurious variations in RV's caused by a faculae using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. 5 and the estimated parameters are presented in Table 2. Once corrected for stellar activity the residuals do not show a systematic component.

Parameter	RV	SN mean RV	SN median RV
β_0	0.0016	0.26	0.60
β_1	$3.75e-05$	0.0068	0.0069
β_2	$7.79e-16$	$1.97e-05$	$3.69e-08$
β_3	0.25	0.0011	0.0011
β_4	0.53	0.21	0.21
R^2	0.9967	0.9571	0.977

Table 2: Evaluation of the linear combination used for correcting the RV's from spurious variations caused by a faculae, according to Eq. 5. Beyond the contrast parameter A and the asymmetry parameter of the CCF which are useful for both the analyses, the width of the CCF is statistically useful to explain spurious variations in RV's caused by a faculae only for the SN analyses. The term of interaction between the asymmetry and the width of the CCF is not helpful to explain part of the variability in the set of RV's. The estimated R^2 show that the proposed correction for stellar activity explains all the spurious variability in RV's.

correlation is observed. Once again, like in the case of the faculae, we see that some parameters of the SN gives stronger correlation than when using the Normal parameters.]]

As before, we corrected the originally RVs by using Eq. 5. The results of the correction are displayed in Fig. 7. Also in this case the proposed correction almost completely addresses the issue when considering the SN or Normal parameters, with values of R^2 larger than 0.98. [[**Xavier:** Looking at Fig. 7, we see that the activity correction proposed here is able to reduce the signal of a spot from a raw RV rms larger than 4.80 $\mathrm{m\,s^{-1}}$down to a rms of 0.38 $\mathrm{m\,s^{-1}}$. In this case the RV rms of the residuals obtained in the case of the Normal parameters is smaller, however,]] [[**Xavier:** REDO WITH FINAL PLOT with a difference of $6\,mm.s^{-1}$, we cannot say that this difference is significant. When comparing the activity correction proposed in this paper with what is commonly used, i.e only a linear dependance with the width and asymmetry of the CCF, we see that our solution is capable of reducing the RV residual rms by a factor of 3.5, which is even more than the factor 2 found in the case of the faculae.

In terms of the significance of the different parameters in Eq. 5, summmurized in Table 3, we observe that the intercept and β_3, corresponding to the width of the CCF, are useless to understand the variation seen in RVs. This is not surprising when looking at the circle shape drawn when plotting the width as a function of the RV in Fig. 6. We see also that the ampltitude parameter, with coefficient β_1 is only useful to explain the RV variation of the spot in the case of the Normal distribution.]]

4.3 Spot and planet

The last simulation presented consists in having a planetary signal influencing the CCF, in addition to the 1% spot modeled before (see Sec. 4.2). [[**Xavier:** The purpose of this example is to check if we are able to disentangle as efficiently the two different sources of variations when using the

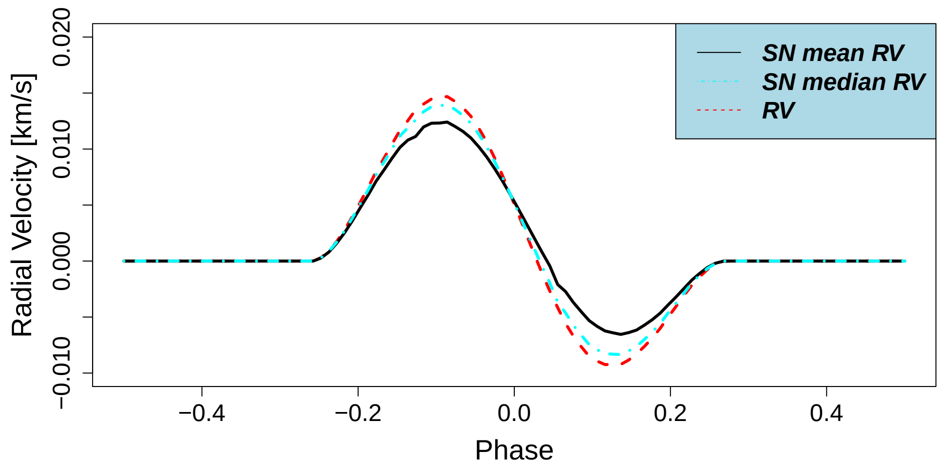

Figure 5: RV's changes as function of the orbital phase in the case in which a spot is present on the photosphere of the star. SN mean RV seems to have the smallest spurious variations caused by the faculae.

Parameter	RV	SN mean RV	SN median RV
β_0	0.47	0.84	0.81
β_1	0.00035	0.064	0.064
β_2	$2e-16$	$2e-16$	$2e-16$
β_3	0.53	0.61	0.61
β_4	$1.11e-07$	$2.85e-08$	$2.87e-08$
R^2	0.9897	0.9914	0.9929

Table 3: Evaluation of the linear combination used for correcting the RV's from spurious variations caused by a spot, according to Eq. 5. All the covariates are statistically useful to explain the variability in RV's causued by a spot except the intercept and the width of the CCF. Concerning the SN analysis, the contrast parameter A is not statistically significant at level 5%. The estimated R^2 show that the proposed correction for stellar activity explains all the spurious variability in RV's.

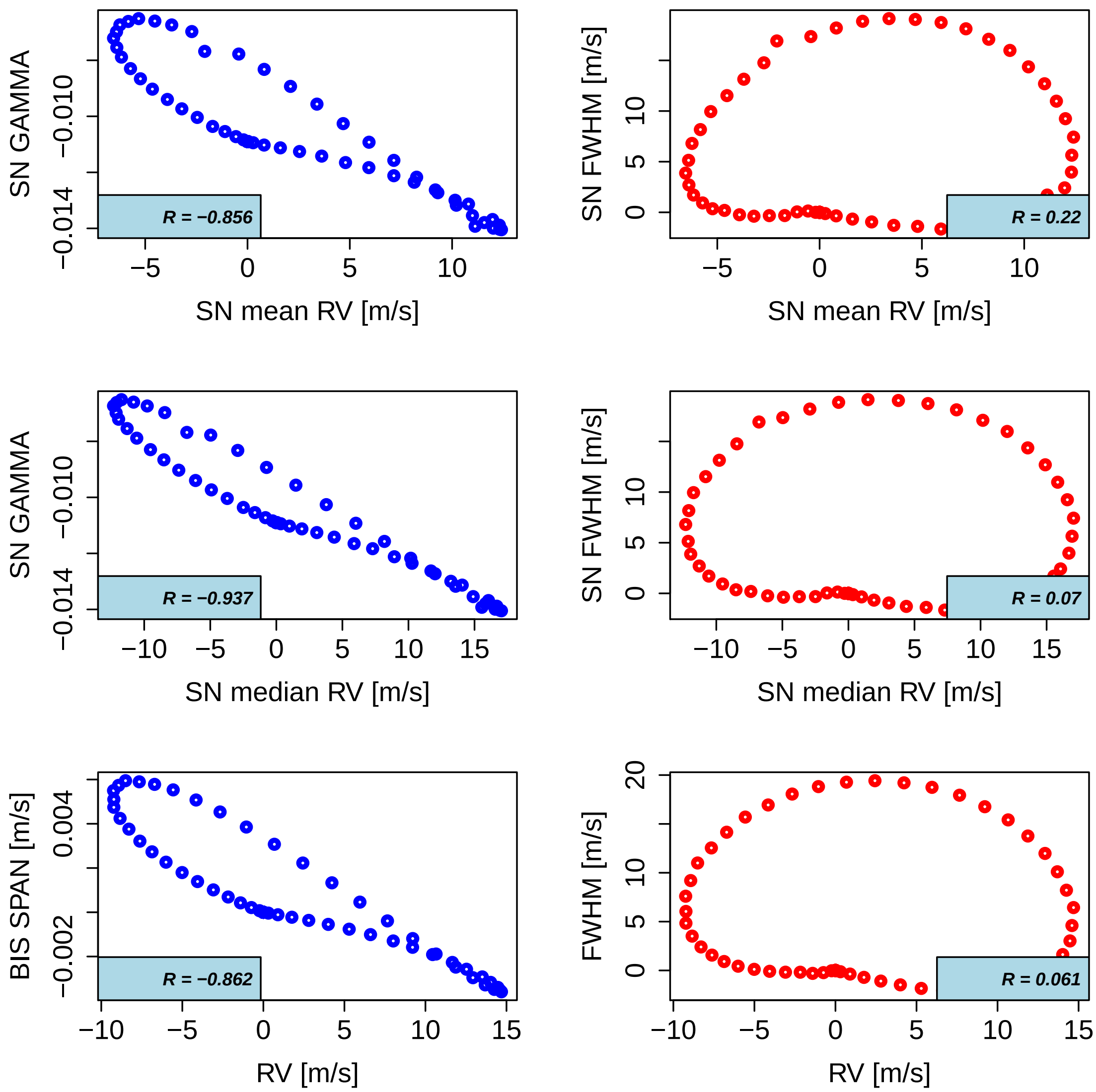

Figure 6: Evaluation of the correlation between the RV's and the asymmetry parameters when a spot is present on the photosphere of the star. In this case only the shape of the CCF changes as the spot moves, producing statistically significative correlations only between the RV's and the asymmetry parameter.

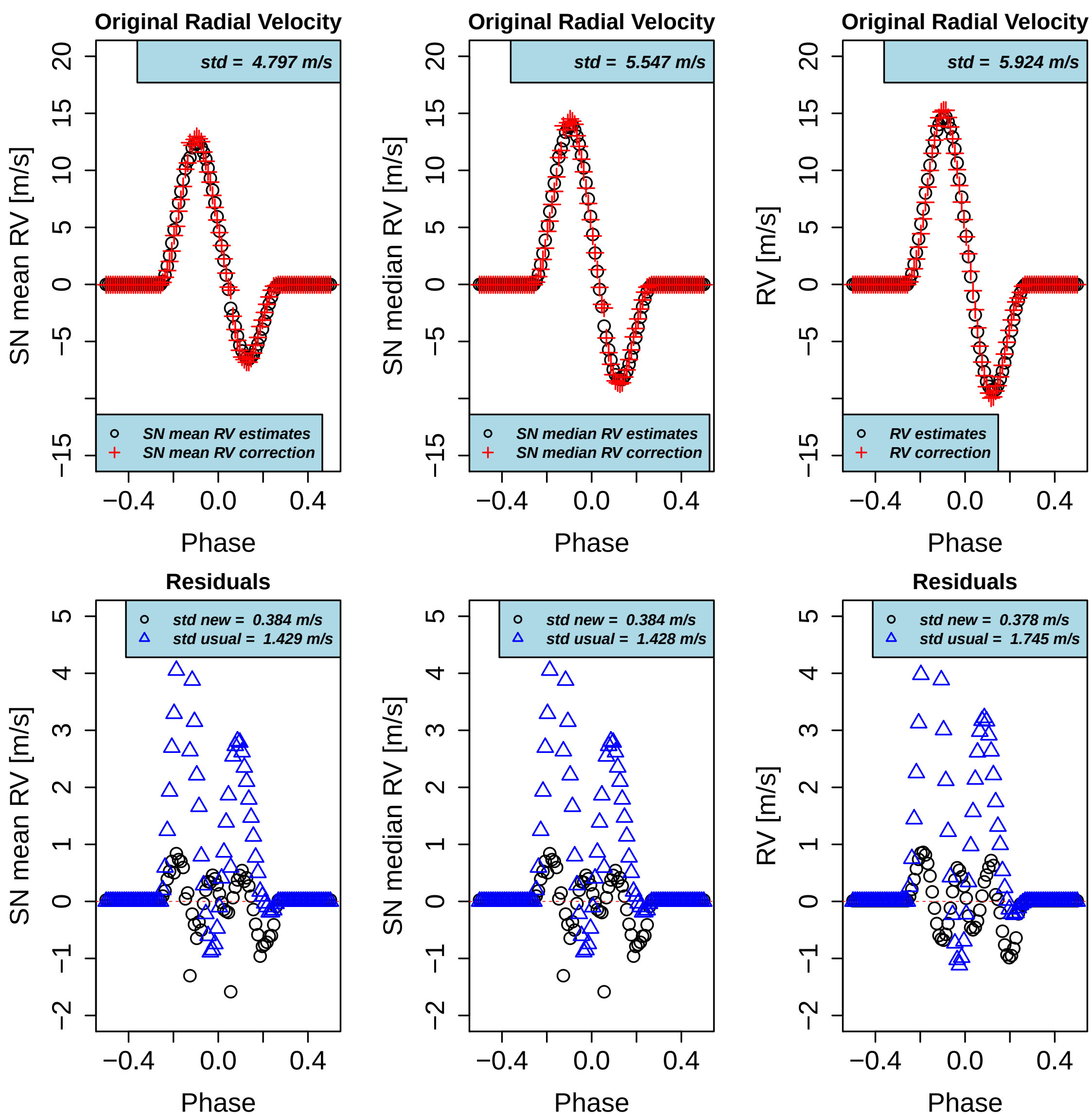

Figure 7: Set of spurious variations in RV's caused by a spot using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. 5 and the estimated parameters are presented in Table 2. Once corrected for stellar activity the residuals do not show a systematic component.

parameters derived using a Normal, or a SN fit the the CCF. In this case, we inject a planet with an semi-amplitude of 10 $\mathrm{m\,s^{-1}}$, with no eccentricity, and with a period corresponding to 1/3rd of the stellar rotational phase.]]

Fig. 8 shows the variation observed in the CCF barycenter parameters. Like in the case of the spot, all barycenter indicators show very similar variations, with SN mean RV showing a slightly smaller amplitude.

In Fig. 9, we show the correlation between the different CCF parameters. [[**Xavier:** Except of seing smaller correlation than in the case of the spot, due to the fact that the planet induces changes in the CCF barycenter without any change in any of the width or asymmetry parameters, the correlation strengths between the asymmetry parameters and the CCF barycenter are in exactly the same order than in the case of the spot: γ–SN median RV $R = -0.84$, BIS SPAN-RV $R = -0.78$ and γ–SN mean RV $R = -0.76$. The variation seen in the width-CCF barycenter space draw a circle like in the case of a spot, therefore, no correlation is observed between those parameters.]]

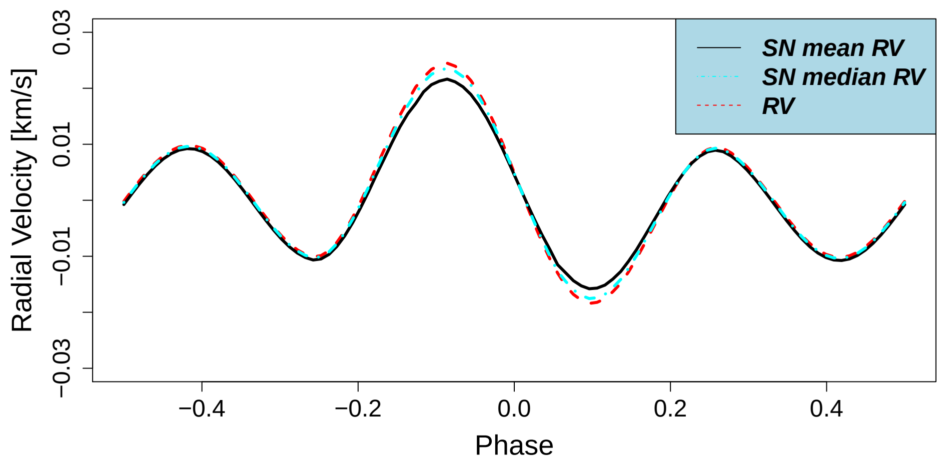

Figure 8: RV's changes as function of the orbital phase in the case in which a spot is present on the photosphere of the star and a planet is injected. N mean RV seems to have the largest variations caused by the combined action of spot and planet.

In order to correct the RVs from the spurious variation caused by the spot, [[**Xavier:** we need to add to our model of activity correction, a signal to take into account the RV variation caused by the injected planet. The observed RV can therefore be modeled by a combination of the activity and the planet signal:]]

$$RV = RV_{\text{stellar activity}} + RV_{\text{planet}}, \tag{6}$$

[[**Xavier:** where $RV_{\text{stellar activity}}$ can be found in Eq. 5, and RV_{planet}, in the case of no eccentricity, can be modelled by the following sinusoidal function:]]

$$RV_{\text{exoplanet}} = K \sin\left(\frac{2\pi}{P}(t - t_0)\right), \tag{7}$$

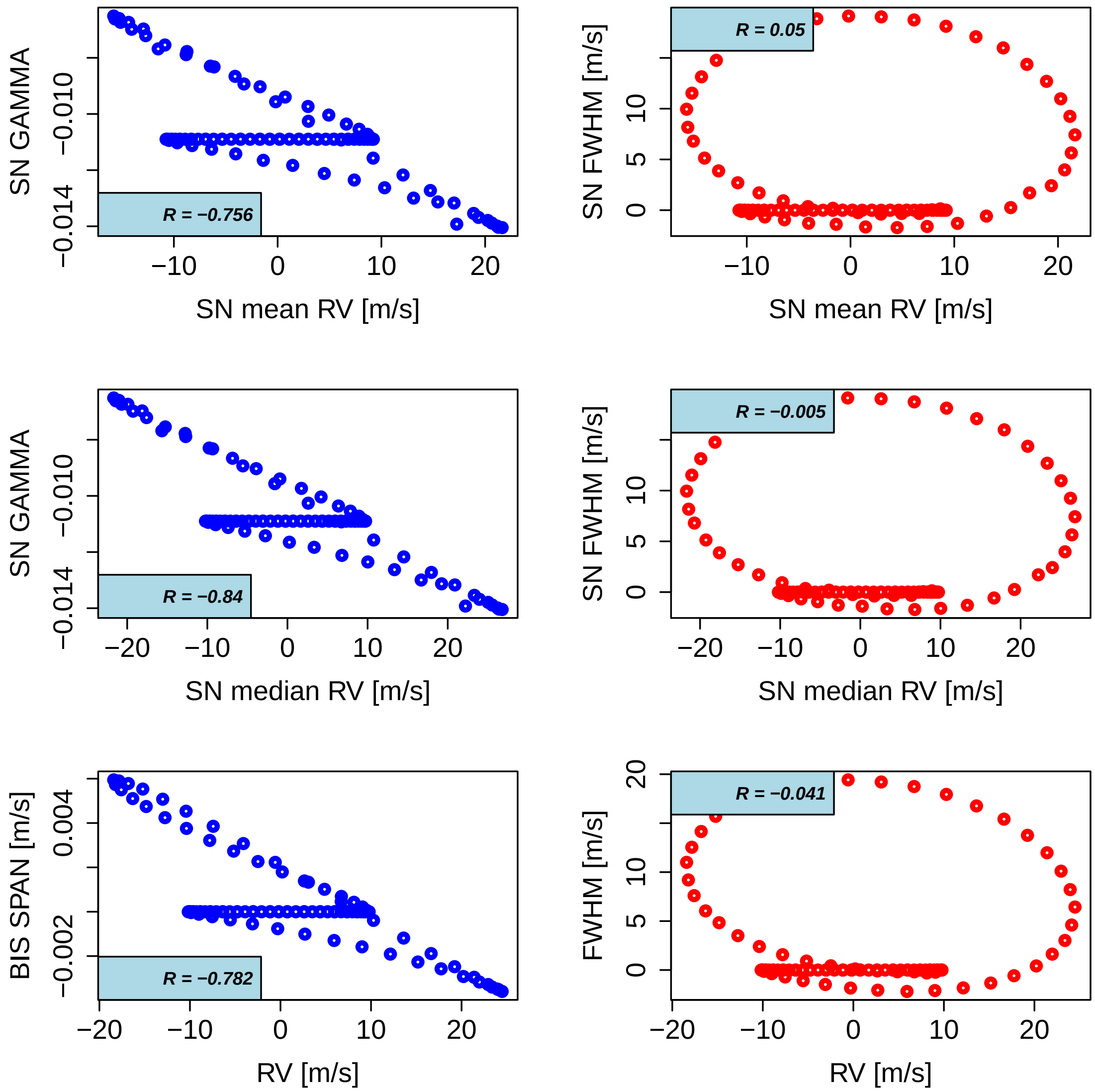

Figure 9: Evaluation of the correlation between the RV's and the asymmetry parameters when a spot is present on the photosphere of the star and a planet is injected. In this case only the shape of the CCF changes as the spot moves, producing statistically significative correlations only between the RVs and the asymmetry parameter. The correlations between the RVs and the width parameter of the CCF is weaker than the previous case that considers only the presence of a spot on the photosphere of the star.

Parameter	RV	SN mean RV	SN median RV
β_0	$7.09e-09$	$3.44e-07$	$4.96e-09$
β_1	0.00105	0.0052	0.0052
β_2	$2e-16$	$2.43e-16$	$2e-16$
β_3	0.79	0.22	0.22
β_4	$6.73e-07$	$7.78e-06$	$7.82e-06$
K	$2e-16$	$2e-16$	$2e-16$
P	$2e-16$	$2e-16$	$2e-16$
t_0	$2e-16$	$2e-16$	$2e-16$
Residuals	$0.61\,\mathrm{m\,s^{-1}}$	$0.60\,\mathrm{m\,s^{-1}}$	$0.60\,\mathrm{m\,s^{-1}}$

Table 4: Evaluation of the linear combination used for correcting the RV's, according to Eq. 6. All the parameters are statistically helpful to address spurious variations in RV's except the FWHM. Concerning the keplerian parameters, the amplitude K that provides relevance about the possibly presence of the exoplanet. Note that since non linear least squares are required, the residual standard error rather than the R^2 is displayed for each case.

where the amplitude K, the orbital period P and the epoch at the periapsis t_0 are three unknown parameters that define the planetary orbit.

The statistical tests conducted on the parameters, whose results are summarized in Table 4, [[**Xavier:** shows that except for the width parameters with coefficient β_3, all the other are significantly useful to explain the RV variation induce by a spot plus a planet.]]

4.4 Conclusions on the simulation study

In this Sec.4, we presented a first implementation of the SN fit to the CCF, using SOAP 2.0 to simulate noiseless CCF affected by stellar activity variation.

[[**Xavier:** is all this discussion really usefull ? Before moving to real cases, where the analyses on five stars are presented, we need to provide further considerations. First of all, looking ad the analyses conducted with SOAP 2.0, it seems that the largest correlation between an asymmetry parameter and a set of RV's happens to be when respectively γ and SN median RV are used. This is a bit surprising, since as the shape of the CCF changes, we expect SN median RV to be more robust than SN mean RV. [[**Umberto:** A possible justification of this ...]]. As second, when searching for stellar activity by deriving the correlation between the set of RV's and either an asymmetry parameter or the width of the CCF, the latter leads to weaker and hence less conclusive results if the active region is a spot. When stellar activity is dominated by faculae, both the shape and the width of the CCF changes as the faculae evolves on the photosphere of the star. Related to these last two considerations, we note that the interaction between the asymmetry and the width of the CCF is useful to explain part of the variability in the RV's if the active region is a spot but not when it is a faculae. The proposed function to correct for stellar activity addressed high level of spurious variations in RV's caused by active regions. In particular, respect to other common linear interpolation, we proposed to use as covariates also the amplitude parameter of the CCF and the interaction between γ and SN FWHM (or BIS SPAN and FWHM). As a consequence of using the interaction between the asymmetry and the width of the CCF, we note that the FWHM (or SN FWHM) becomes statistically not significant, while this is not the case if the interaction term is not involved in the linear regression. Finally, the correlations involving the common indicators (i.e. RV, FWHM and BIS SPAN) are systematically weaker than the correlations obtained by fitting the SN to the CCF, suggesting that this density could be helpful when searching for active regions. We recall moreover that all the quantities needed for conducting the analyses of the CCF are directly available by just fitting the SN.]]

5 Real data application

In this Section we present the analyses conducted on Alpha Centauri b, comparing the result of fitting a CCF using the SN density defined in Sec. 2.1 with the approach based on fitting a Normal density to retrieve the RV and width of the CCF and calculating the bisector to derive the asymmetry parameter BIS SPAN. Four other stars have been analyzed with the proposed method and details can be found in the Appendix A. For all the stars that have been considered in the present work, [[**Xavier:** we selected CCFs that were derived from spectra that had a SNR at 550 nm larger than 10.]]

5.1 Alpha Centauri B

We analysed the 1808 RV observations of Alpha Centauri B taken in 2010 by the HARPS spectrograph. Note that more observation have been done during this year, however we selected here only the data that are not significantly affected by contamination from Alpha Cen A (see Dumusque et al. 2012). [[**Xavier:** We choose these observations as their represent probably the best sampled and most precise RV data set showing a strong solar-like activity signal (Thompson et al. 2017; Dumusque et al. 2012).]]

We begin the analyses by evaluating the correlation between γ and the BIS SPAN. [[**Xavier:** As γ has no units, this will allow us to compare the amplitude of the activity signals seen in γ and in BIS SPAN, by using the slope of this correlation as a scaling factor]]. In Fig. 10, we see that the relationship between γ and the BIS SPAN is linear, with a slope equal to 0.72 and a strong Pearson correlation coefficient of $R = 0.95$.

Fig. 11 shows the comparison between the RV's retrieved using the SN shape and the ones obtained with the Normal shape. It is possible to appreciate the presence of a strong stellar activity signal, as expected (Dumusque et al. 2012; Thompson et al. 2017). When using SN mean RV, it is possible to observe more variations than the ones measured by the Normal fitting. This happens because the mean of the SN is more sensitive to stellar activity. In fact, because the SN includes an asymmetry parameter, SN mean RV gets more shifted in the direction of the asymmetry induced by stellar activity. On the other hand, when using SN median RV, smaller variations in RV are caused by changes in the asymmetry of the CCF, because this second location parameter is a more robust

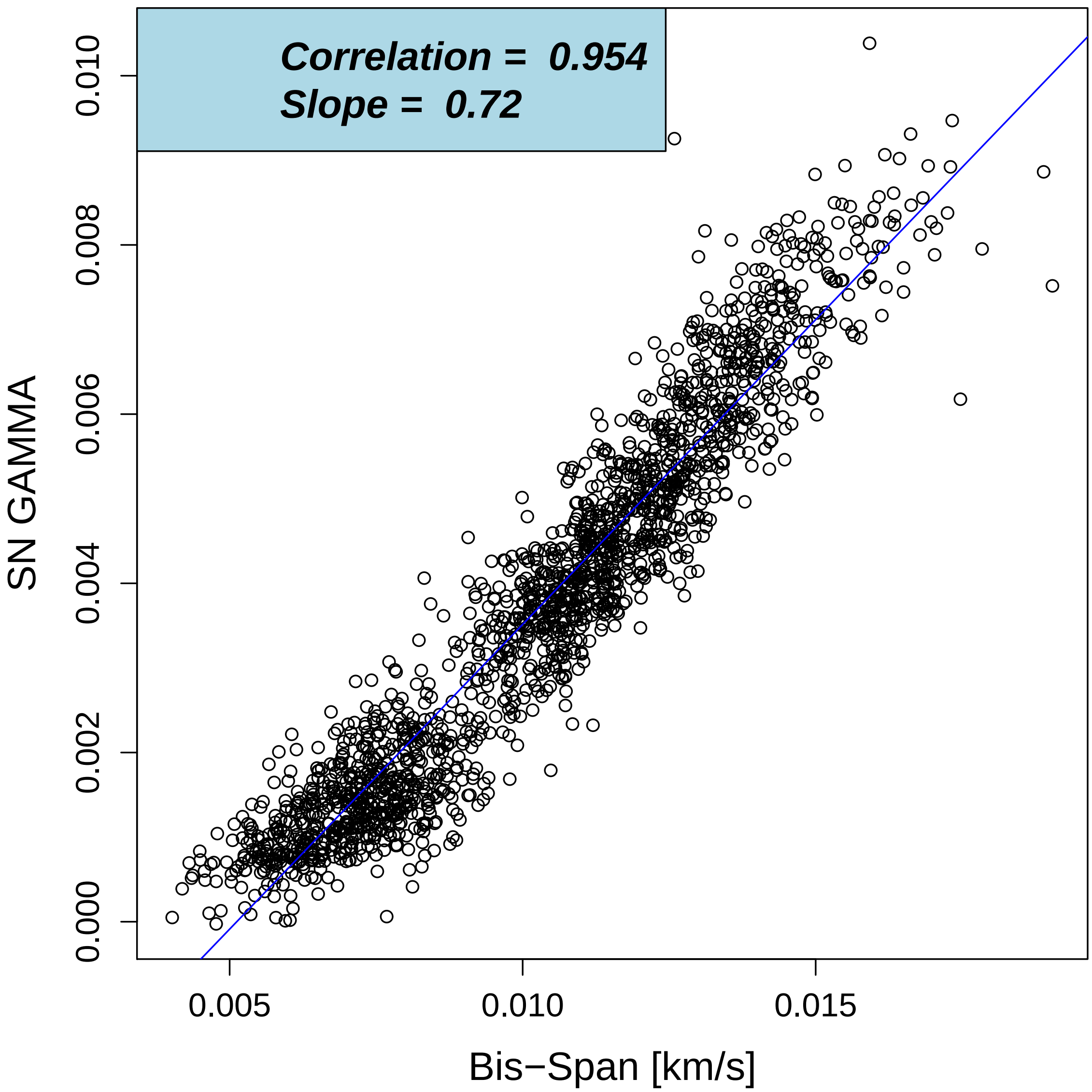

Figure 10: Correlation between γ and the BIS SPAN for Alpha Centauri B. Because γ is adimensional, retrieving the slope between γ and the BIS SPAN, which is expressed in $\mathrm{km\,s^{-1}}$, allows us to provide physical meaning to γ.

indicator than the mean. The bottom plot of Fig. 11 captures this aspect. Both indicators can be used to capture and summarise the different information available in the CCF, as will be shown in the remainder of this work.

Similar to Figueira et al. (2013), we compare the correlation between the different activity indicators and the RV's of the star in Fig. 12. The correlation between γ and SN mean RV and the correlation between γ and SN median RV are much stronger than the correlations calculated between the other asymmetry parameters and their corresponding RV's. In particular the correlation between γ and SN mean RV is almost twice the correlation between the other asymmetry parameters and their corresponding RV's.

Because the median is a more robust index than the mean, the correlation between γ and SN median RV is not as large as the correlation between γ and SN mean RV, but it is nonetheless 1.5 times larger than the correlation between the other common asymmetry parameters and their corresponding RV's. In other words, changes in the asymmetry of the CCF are better captured when using the SN mean RV. The correlation between FWHM and the RV's, either by using SN mean RV or SN median RV, is as well stronger when fitting a SN density rather than a Normal. All the correlations are statistically different from 0. Recalling the analyses presented in Sec. 4, we could infer that Alpha Centauri b is dominated by faculae, because the correlations between the RV's and the width of the CCF are strong (in particular the correlation between SN mean RV and SN FWHM is 0.817).

Using Eq. 5, we provide a new set of RV's corrected from stellar activity. The results are shown in Figure 13. We see that, once corrected for stellar activity, the residuals in the Normal and SN analysis are comparable. However, we note that when using SN mean RV, the correction is more important. In fact, the comparison of R^2 shows that the SN fit accounts for a higher percentage of variability caused by stellar activity (i.e. spurious variations in RV's). We note also that the BIS SPAN is not helpful to address part of the spurious variability caused by stellar activity, while the opposite conclusion has reached when evaluating the p-value related to γ.

Both the proposed indicators coming from the SN density have advantages and limits: SN mean RV better catches changes in the asymmetry of the CCF but the resulting set of RV's ends up being contaminated by those spurious shifts caused by stellar activity that have been shortly presented in Sec. 1. When using SN median RV, the final set of RV's is less affected by those spurious shifts caused by stellar activity, but at the same time this indicator is not able to catch as well as SN mean RV changes in the shape and in the width of the CCF. Once corrected from stellar activity using Eq. 5, the results are comparable. Anyway, both SN mean RV and SN median RV are useful to catch different aspects of the CCF and our suggestion is to use SN mean RV when interested in retrieving information about changes in the shape and/or the width of the CCF. In order to provide a set of RV's containing the smallest amount of spurious contamination imputable to stellar activity (i.e. before to run Eq. 5), our suggestion is to use instead SN median RV.

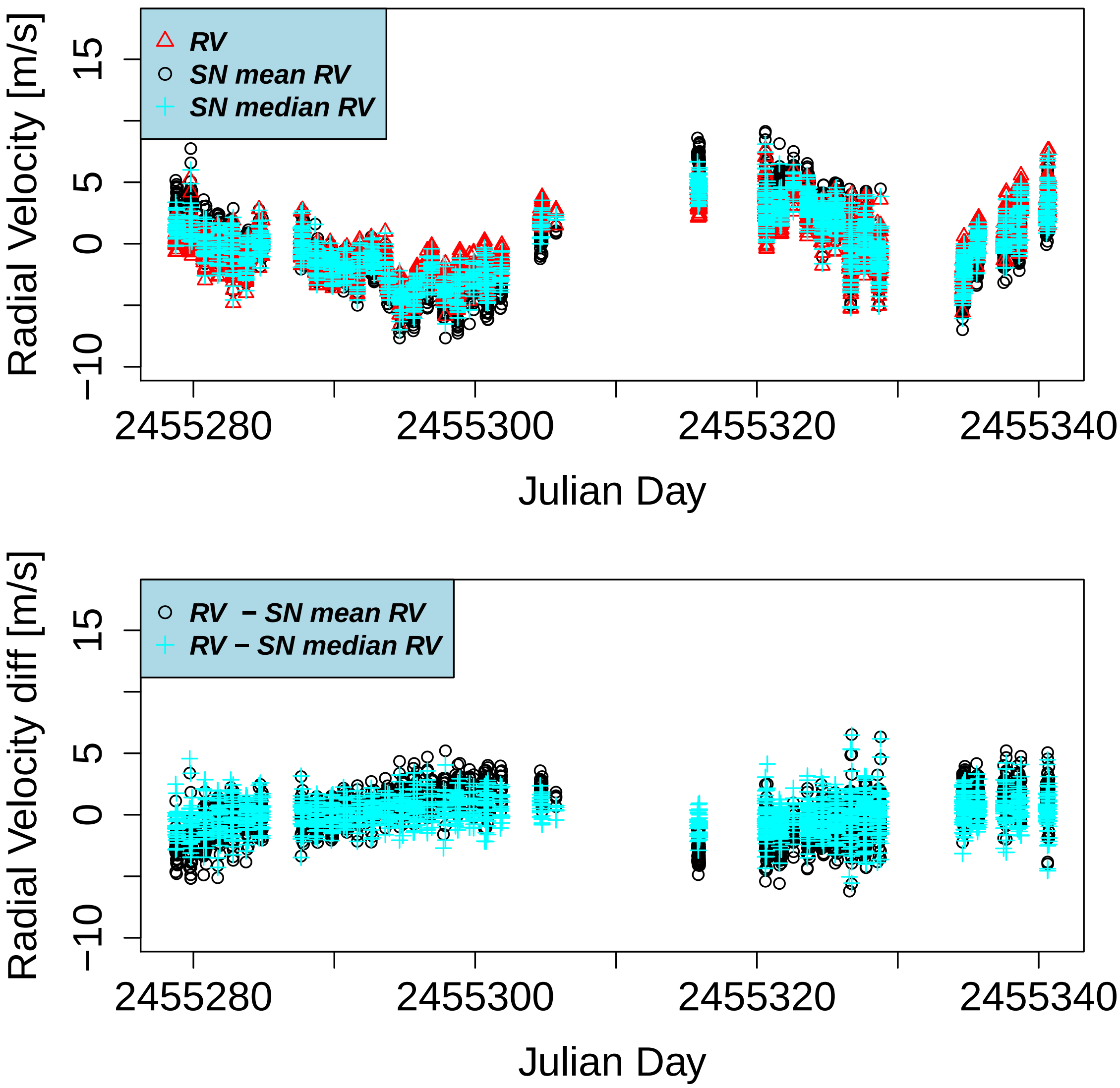

Figure 11: (top) RV's as function of Julian Day for Alpha Centauri b. The RV's are retrieved using the mean of the Normal (red triangles), SN mean RV (black circles), SN median RV (cyan crosses). (bottom) RV differences between Normal RV and SN mean RV (black circles) and between Normal RV and a SN median RV (cyan crosses).

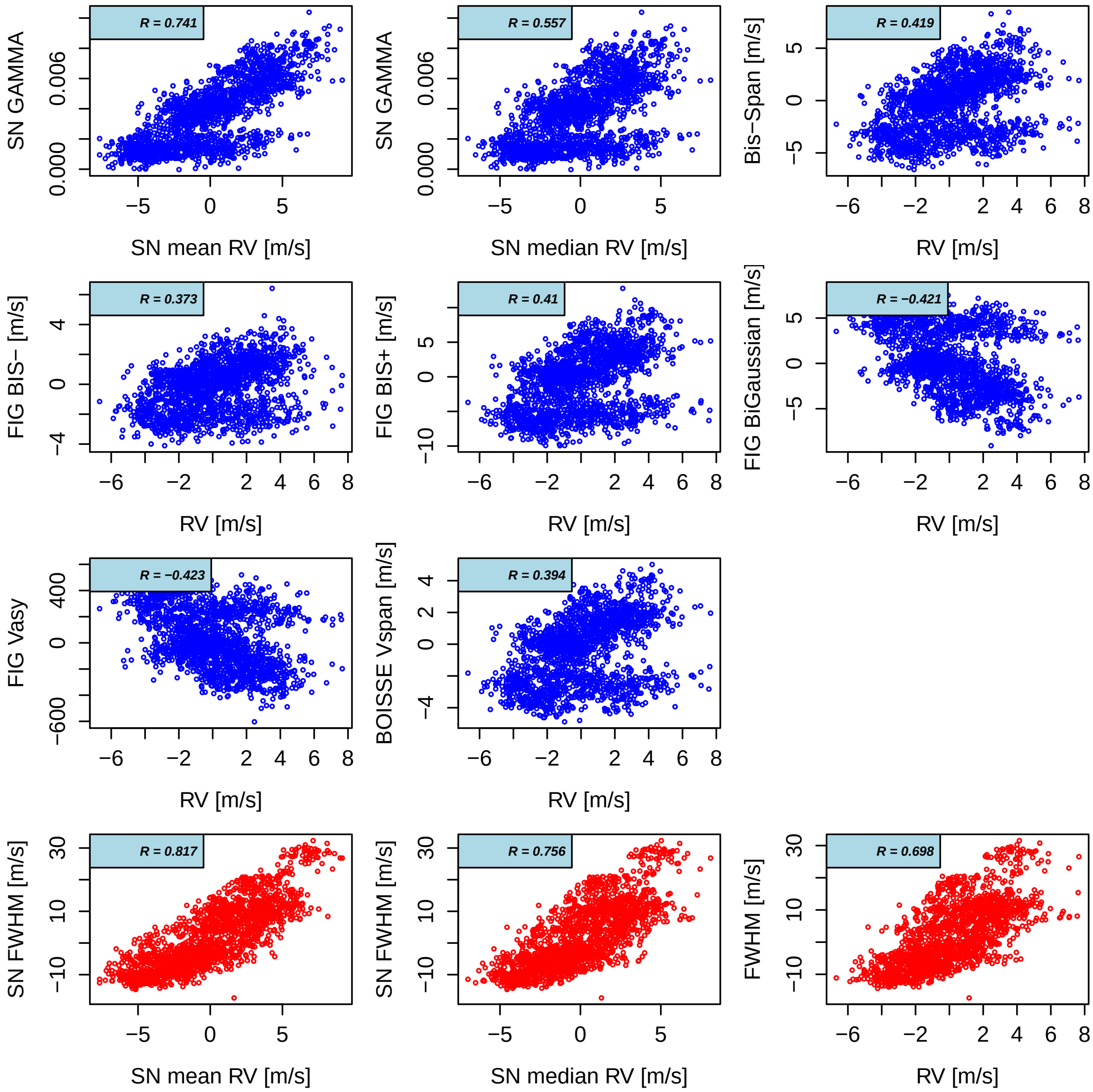

Figure 12: Correlation between the asymmetry parameters and the RVs for Alpha Centauri B. The last three plots show the correlation between the FWHM and the RVs for Alpha Centauri B using respectively the SN (SN mean RV and SN median RV) and the Normal fits. The correlations are always stronger when using parameters derived from the SN fit than the Normal one. The p–values associated with each *R* is statistically different from 0.

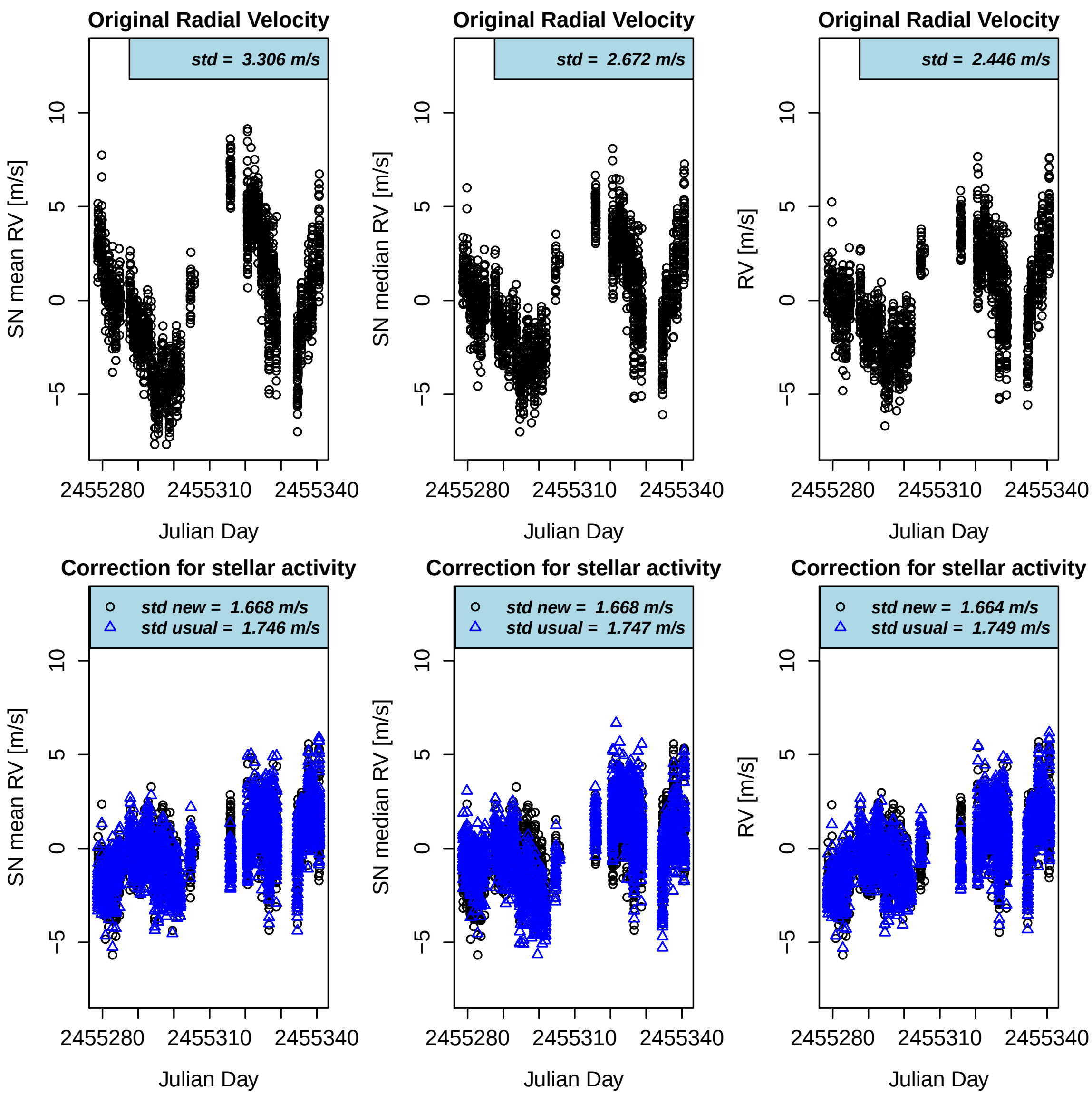

Figure 13: Set of RV's for Alpha Centauri B using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. 5. Once corrected for stellar activity, the residuals in the Normal and SN analyses are comparable.

Parameter	RV	SN mean RV	SN median RV
β_0	0.49	0.90	0.027
β_1	$2.22e-16$	$2.22e-16$	$2.22e-16$
β_2	0.33	$2.22e-16$	$1.23e-11$
β_3	$2.22e-16$	$2.22e-16$	$2.22e-16$
β_4	$2.22e-16$	$2.22e-16$	$2.22e-16$
R^2	0.57	0.78	0.66

Table 5: Evaluation of the linear combination used for correcting the RV's, according to Eq. 5. Concerning the Normal fit, all the parameters but the intercept and the BIS SPAN are useful in explaining variations in RV's of the star that can be caused by stellar activity. For the SN fit we note that the parameter related to γ is highly significant to address part of the spurious variations in RV's caused by stellar activity. The evaluation of the R^2 shows that the proposed linear combination better explains variations in RV's due to stellar activity coming from the SN analysis that uses SN mean RV.

5.2 Doppler-shift added to Alpha Centauri B

We also consider a real-data example using HARPS spectra from the star Alpha Centauri B with an imputed Doppler-shift added...

[[**Umberto:** to be done]][[**Xavier:** to plot K vs P for Normal, SN mean, SN median]]

In the next Section we further motivate the reasons to define the RV's derived by the CCF by calculating SN median RV. In order to do that, we retrieve the standard errors associated with SN mean RV, SN median RV and RV.

6 Estimation of standard errors for the CCF parameters

In this Section, we perform a bootstrap analysis (Davison and Hinkley 1997; Efron and Tibshirani 1994) in order to retrieve the standard errors associated to SN mean RV, SN median RV, RV, SN FWHM, FWHM, γ and BIS SPAN. Because a CCF is obtained from a cross-correlation, each point in a CCF is correlated with each other. Therefore, we cannot do a bootstrap analysis on perturbing independently each CCF point with a Normal density scaled to the error of each given point. A detailed discussions of the methods nowadays available to resampling in situations with dependent data structures is available in Lahiri (2013). All the bootstrap methods that deal with dependant data structures rely on the so called Block Bootstrap methods, originally introduced by Wilks (1997). In our particular case, since each point in a CCF is correlated with each other, we bootstrap a hundred times the stellar spectrum given the photon-noise error of each wavelength and calculate for each realization a new CCF. We then fit a Normal or a SN density to each of these CCF's and then calculate the standard deviations of the density for the location parameters (RV, SN mean RV or SN median RV), the width parameters (SN FWHM or FWHM) and the asymmetry parameters (γ or BIS SPAN).

6.1 Estimation of standard errors for the CCF parameters for the simulation study

We start by calculating the standard errors of the parameters retrieved in Sec. 4, where with SOAP we produced CCF's presenting spurious variations in RV's caused by a faculae or by a spot. In the third and final case we considered, beyond the spot, a planetary signal that produces pure Doppler-shifts in the CCF's.

Fig. 14 shows the results of the bootstrap analysis performed when a faculae is present on the photosphere of the star. The series of three plots in the top of Fig. 14 shows the different errors for the RV's, defined as RV (red triangles), SN mean RV (black circles) or SN median RV (cyan crosses), the width of the CCF, defined as FWHM (red triangles) or SN FWHM (black circles) and the asymmetry of the CCF, defined as γ (black circles) or BIS SPAN (red triangles). In the series of three plots in the bottom of Fig. 14 we show the ratio between errors associated to the parameters derived from the bootstrap analysis fitting the SN and the errors associated to the parameters derived from the bootstrap analysis fitting the Normal density. We used the same notation also for the other two cases, where respectively a spot is present on the photosphere of the star (Fig. 15) and a spot and a planet are introducing both spurious and pure variations in the RV's (Fig. 16). Concerning the standard errors related to the RV's if a faculae is creating spurious signals, the ratio between the RV error measured by the bootstrap using the SN and Normal fitting is 1.4, when using SN mean RV and 0.8 when using SN median RV. By using SN median RV we get standard errors 20% smaller than using the Normal fit and its corresponding mean. Regarding the errors in width of the CCF, we see that the bootstrap analysis for the Normal and the SN are comparable. Therefore, the precision in the width of the CCF is the comparable if we fit a Normal or a SN to the CCF. Finally, for the errors in evaluating the asymmetry of the CCF, we see that, when fitting the SN to the CCF, the asymmetry errors are 20% smaller. Therefore, the SN fit gives a better precision in CCF asymmetry than what can be reached using BIS SPAN.

Fig. 15 shows the results of the bootstrap analysis performed when a spot is present on the photosphere of the star. The series of plots follows the specifications outlined for the previous case. Concerning the standard errors related to the RV's, the ratio between the RV error measured by the bootstrap using the SN and Normal fitting is 1.3 when using SN mean RV and 0.8 when using SN median RV. Regarding the errors in width of the CCF, we see that the bootstrap analysis for the Normal and the SN are comparable. Therefore, the precision in the width of the CCF is the comparable if we fit a Normal or a SN to the CCF. Finally, for the errors in evaluating the asymmetry of the CCF, we see that, when fitting the SN to the CCF, the asymmetry errors are 20% smaller.

Fig. 16 shows the results of the bootstrap analysis performed when a spot is present on the photosphere of the star. The series of plots follows the specifications outlined for the previous two cases. The conclusions are comparable to the case in which only a spot is present on the photosphere of the star. The ratio between the RV error measured by the bootstrap using the SN and Normal fitting is 1.3 when using SN mean RV and 0.8 when using SN median RV. The errors in width of the CCF are comparable and the errors in evaluating the asymmetry of the CCF are 15% smaller when using the asymmetry parameter γ of the SN.

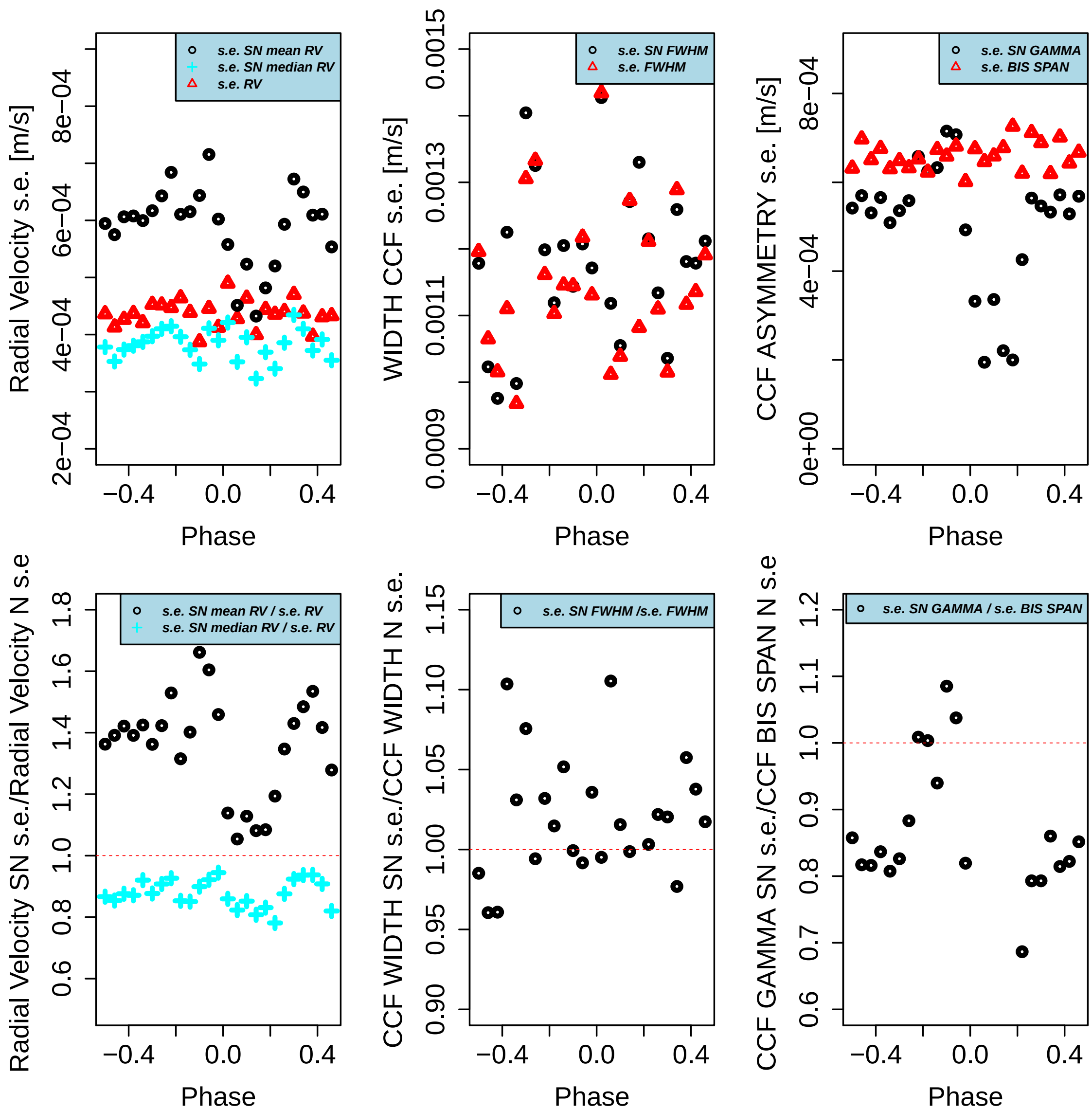

Figure 14: Faculae Case. Comparison between the standard errors using the bootstrap analysis for the RV's, the FWHM and the asymmetry parameter. When using SN mean RV (black circles), the standard errors are in average 40% larger than the standard errors retrieved fitting a Normal (red triangles). However, if using SN median RV (cyan crosses), the standard errors are in average 20% smaller than the standard errors coming from the Normal fit. To use as asymmetry parameter γ of the SN leads to standard errors in average 20% smaller than the standard errors related to the BIS SPAN. [[**Umberto:** explain what happens for those CCF 15 to 19 where s.e. decrease.]] Note that for the asymmetry, the error in BIS SPAN is in $\mathrm{km\,s^{-1}}$. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

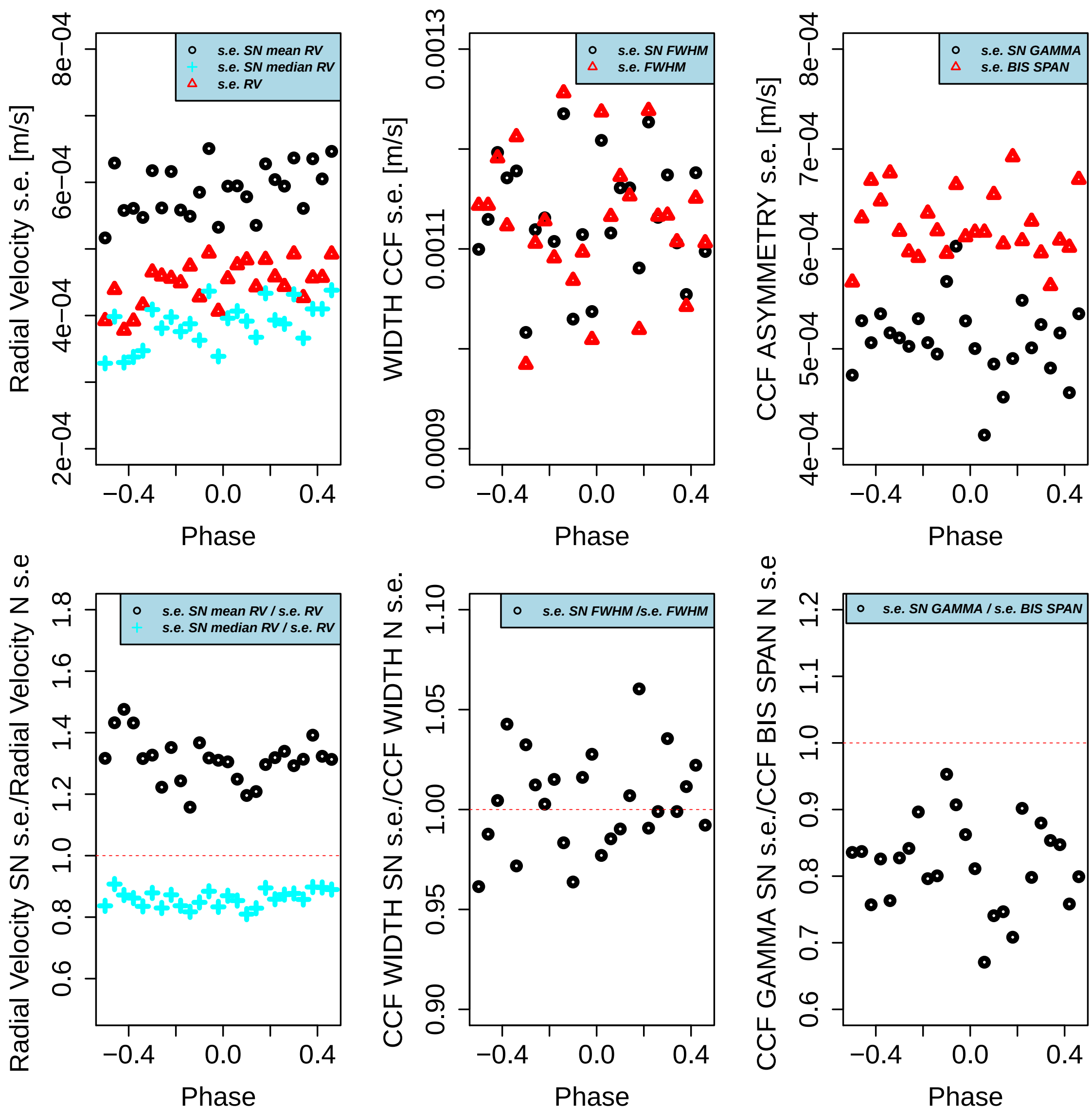

Figure 15: Spot case. Comparison between the standard errors using the bootstrap analysis for the RV's, the FWHM and the asymmetry parameter. When using SN mean RV (black circles), the standard errors are in average 30% larger than the standard errors retrieved fitting a Normal (red triangles). However, if using SN median RV (cyan crosses), the standard errors are in average 20% smaller than the standard errors coming from the Normal fit. To use as asymmetry parameter γ of the SN leads to standard errors in average 20% smaller than the standard errors related to the BIS SPAN. Note that for the asymmetry, the error in BIS SPAN is in $\mathrm{km\,s^{-1}}$. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

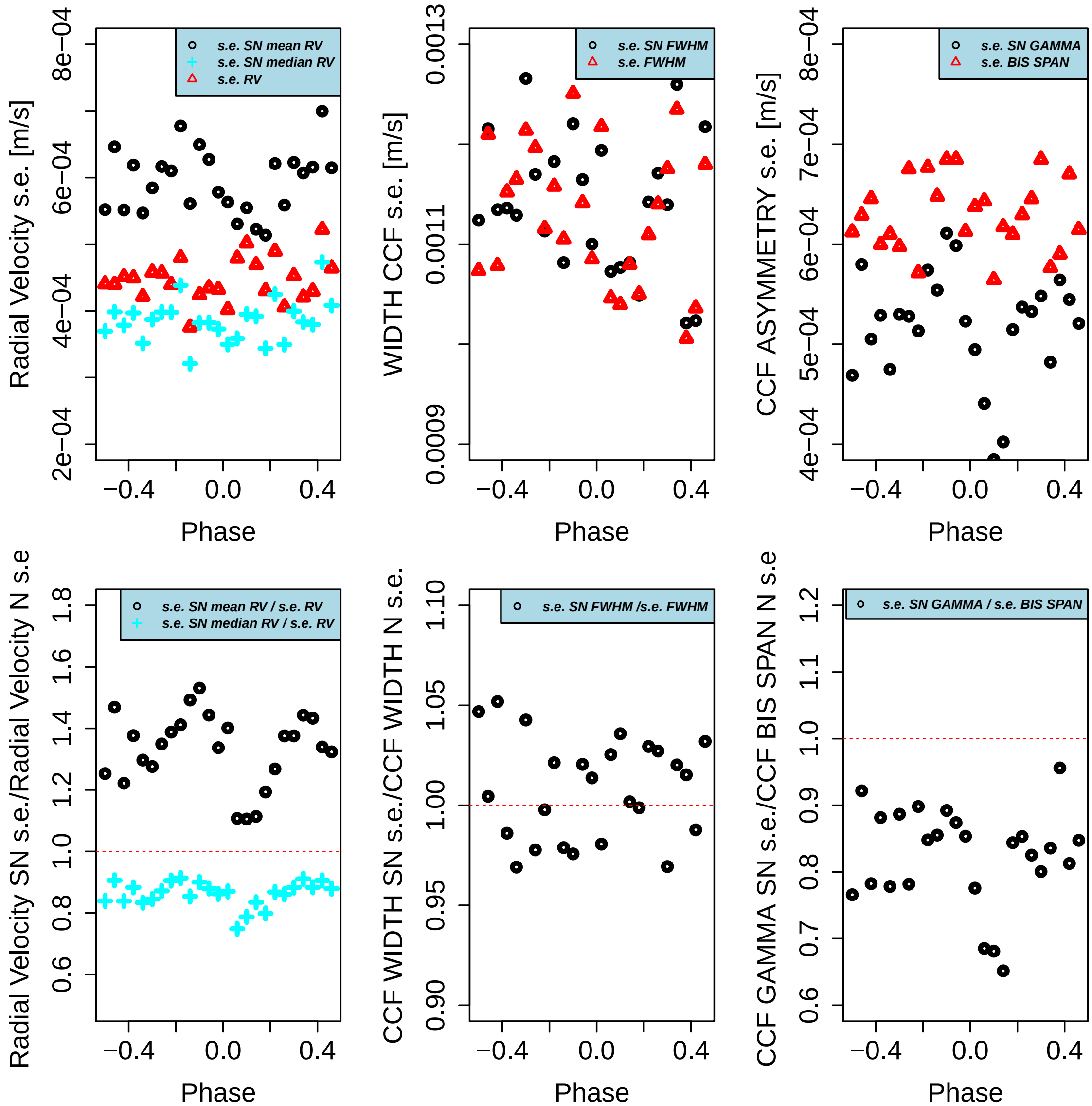

Figure 16: Spot and Planet case. Comparison between the standard errors using the bootstrap analysis for the RVs, the FWHM and the asymmetry parameter. When using SN mean RV (black circles), the standard errors are in average 30% larger than the standard errors retrieved fitting a Normal (red triangles). However, if using SN median RV (cyan crosses), the standard errors are in average 20% smaller than the standard errors coming from the Normal fit. To use as asymmetry parameter γ of the SN leads to standard errors in average 15% smaller than the standard errors related to the BIS SPAN. Note that for the asymmetry, the error in BIS SPAN is in $\mathrm{km\,s^{-1}}$. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

6.2 Estimation of standard errors for the CCF parameters for real stars

In the top plots of Fig. 17 we show the different errors for the RVs, either defined as RV (red triangles), SN mean RV (black circles) or SN median RV (cyan crosses), the width and the asymmetry of the CCFs for three star, HD215152, HD192310 and Corot-7, that are all at different SNR levels. The parameter SN50 corresponds to the SNR in order 50, which defines a wavelength of 550 nm. In the bottom plots, we show the ratio between the parameters derived from the bootstrap analysis fitting the SN and the parameters derived from the bootstrap analysis fitting the Normal density. We first see that the errors on the CCF parameters only depends on the SNR and do not depend on the spectral type. This is true if the spectral type are not too different though, like here where we show the results for G and K dwarfs.

Concerning the standard errors related to the RVs, the ratio between the RV error measured by the bootstrap using the SN and Normal fitting is 1.6 when using SN mean RV and 0.9 when using SN median RV. In other words, by using SN median RV as parameter that defines the radial velocity of the star given a CCF, we get standard errors 10% smaller than using the Normal fit and its corresponding mean. This result is consistent with what we observed with the simulation from SOAP presented in Sec. 4.

Regarding the errors in width of the CCF, we see that the bootstrap analysis for the Normal and the SN are comparable. Therefore, the precision in the width of the CCF is the comparable if we fit a Normal or a SN to the CCF.

Finally, for the errors in evaluating the asymmetry of the CCF, we see that, when fitting the SN to the CCF, the asymmetry errors are 15% smaller. Therefore, the SN fit gives a better precision in CCF asymmetry than what can be reached using BIS SPAN. We recall moreover that, using the SN, all parameters are automatically retrieved in 1 single step, while in the common approach the RV and the BIS SPAN are calculated separately.

7 Discussion

An analysis of the CCF residuals after fitting a Normal or SN density shows that the SN is a slightly better model to explain the shape of the CCF. This comes from the fact that CCF's present a natural asymmetry due the convective blueshift.

We tested at first our assumptions by using simulated CCF's retrieved using the software SOAP 2.0. We then compared for five real stars the difference between the RV's (defined as mean of a Normal, mean of the SN or median of the SN), FWHM and asymmetry (BIS SPAN in the Normal case and γ in the SN case). The γ parameter is linearly dependent on the BIS SPAN, with always a strong correlation coefficient. The slope of this linear correlation changes depending on the studied star. This is probably because the spectral type is different, therefore the effects from stellar activity are different.

When using as parameter for the RV the mean of the SN, the standard errors are in average 60% larger than the standard errors retrieved fitting a Normal. However, once the RV is defined as the median of the SN (cyan crosses), the standard errors are in average 10% smaller than the standard

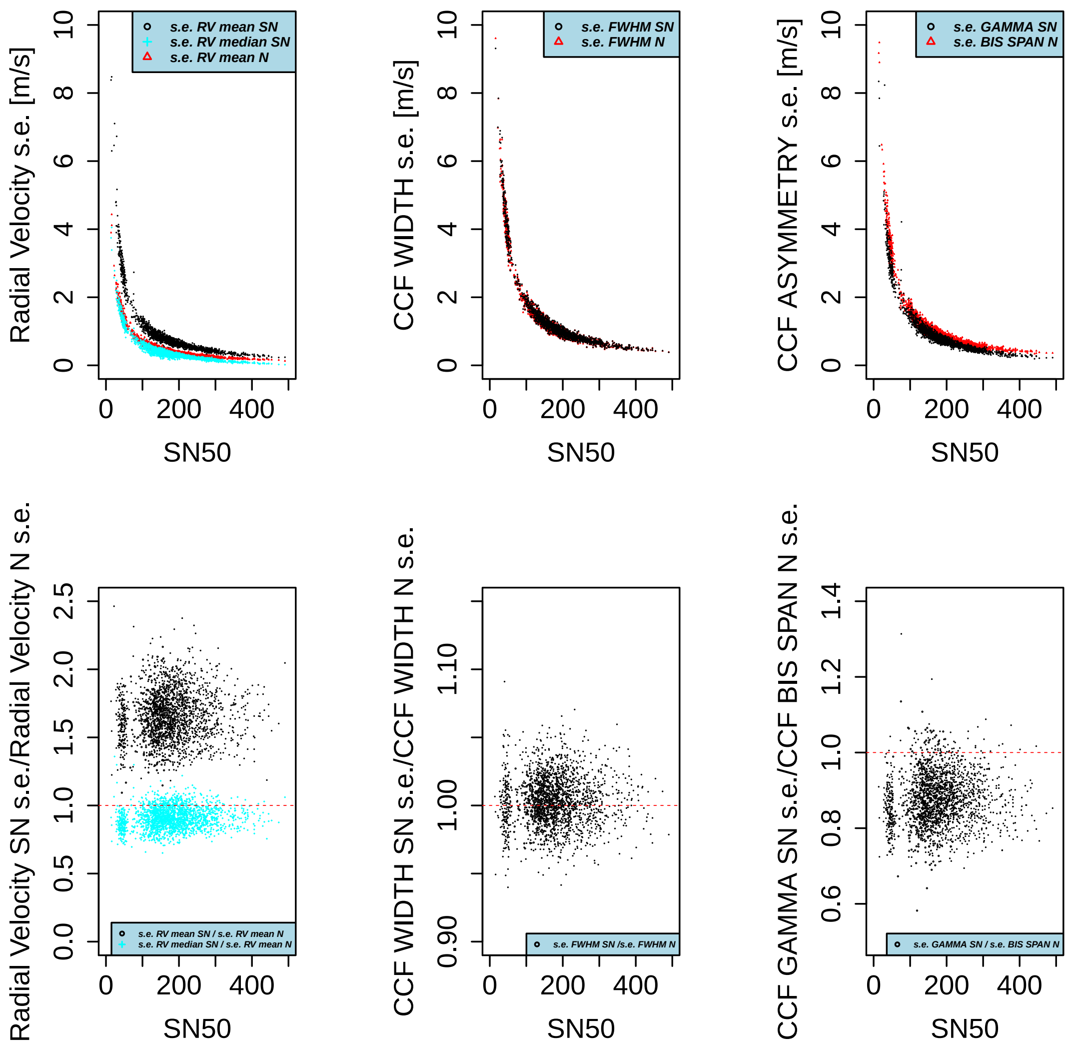

Figure 17: Comparison between the standard errors using the bootstrap analysis for the RVs, the FWHM and the asymmetry parameter. When using SN mean RV (black circles), the standard errors are in average 60% larger than the standard errors retrieved fitting a Normal (red triangles). However, if using SN median RV (cyan crosses), the standard errors are in average 10% smaller than the standard errors coming from the Normal fit. To use as asymmetry parameter γ of the SN leads to standard errors in average 15% smaller than the standard errors related to the BIS SPAN. Note that for the asymmetry, the error in BIS SPAN is in $\mathrm{km\,s^{-1}}$. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

errors coming from the Normal fit. When looking at the correlation between the asymmetry and width parameters of the CCF (FWHM and BIS SPAN or the alternative indicators in Figueira et al. (2013) in the Normal case, and SN FWHM and γ in the SN case) with respect to the RV's (RV's in the Normal case or SN RV's in the SN case), we observe that the correlations are always stronger for the parameters of the SN. Therefore, the SN parameters are more sensitive to activity. In the case of Tau Ceti, which is at very low activity level, we find a significant correlation of 0.322 between γ and SN mean RV, while for all the other asymmetric parameterization, BIS SPAN or the alternative indicators in Figueira et al. (2013), the correlations are weaker with a maximum of 0.225.

8 Conclusion

In this paper we introduced a novel approach based on the Skew Normal (SN) density to retrieve RV's and shape variations in the CCF of stars. When searching for small-mass exoplanets using the RV technique, it is essential to understand variation of the shape of the CCF, which is a proxy for stellar activity effects. The standard approach consist at first to fit a Normal density to the CCF to get the RV and FWHM, defined as the mean and the FWHM of the Normal density, and then to measure the asymmetry of the CCF by calculating the BIS SPAN. FWHM and BIS SPAN give us information on the line shape that are used to probe stellar activity signals.

In this paper we propose to conduct the analysis fitting a SN density to the CCF. Since the CCF presents a natural asymmetry due the convective blueshift, the SN density can in principle better catch these aspects respect the Normal fit. On top of that, by using the SN density to fit CCF's, we can measure simultaneously the RV of the star, the width and the asymmetry of the CCF.

Starting from the simulation environment SOAP and then moving to real stars, we showed that using the SN density to fit CCF's leads to a significant improvement to probing stellar activity. While for the Normal density mean and median are equivalent, using the SN fit different location parameters can be tested. While the median of the SN is a more robust statistic respect variations in the shape of the CCF, the mean of the SN is more sensible to changes in the asymmetry of the CCF. We suggest to use as parameter that defines the RV of the star the median of the SN, since the standard errors related to this parameter working with CCF's from real stars are on average 10% smaller than the standard errors retrieved using the Normal density. To evaluate changes in the asymmetry of the CCF, we suggest to use the mean of the SN. The correlation between SN mean RV and SN FWHM and the correlation between SN mean RV and γ (the asymmetry parameter of the SN) are much stronger than the correlations between the equivalent parameters derived using a Normal fit (RV, FWHM and BIS SPAN or the asymmetric parameters described in Figueira et al. (2013)). The precision on the asymmetry measured by γ is greater than the one on BIS SPAN by $\sim$15%. Therefore when searching for rotational periods in the data, or applying Gaussian Processes to account for stellar activity signals, the SN parameters should be used.

Because of stellar activity the estimated RV's are contaminated by spurious variations. We propose to use a function that corrects from stellar activity that beyond the width and the asymmetry parameters of the CCF includes also the contrast parameter A and the fourth parameter defined as the interaction between width and the asymmetry parameters of the CCF. We found these new two parameter helpful to explain part of the spurious variations in RV's caused by stellar activity.

Finally, we also encourage the use of bootstrapping to estimate more realistic errors on the different parameters of the Normal or SN fitted to the CCF, mainly in the low SNR regime where a gain of 50% can be reached. This takes significantly more time, but note that 100 bootstrapped dataset are enough to get a good estimation of errors.

9 Acknowledgements

We are grateful to all technical and scientific collaborators of the HARPS Consortium, ESO Headquarters and ESO La Silla who have contributed with their extraordinary passion and valuable work to the success of the HARPS project. XD is grateful to the Society in Science–The Branco Weiss Fellowship for its financial support.

A Appendix

In this Appendix we present the analyses conducted on other 4 stars: HD192310, HD10700, HD215152 and finally Corot-7.

[[**Umberto:** Add further information about the stars here presented.]][[**Xavier:**]]

Table 6 summarizes the results obtained by the SN fit and the some of the results based on the Normal fit. The results are all consistent with the conclusions derived by the analyses on Alpha Centauri b. The correlation between γ and SN mean RV is stronger than the correlation between the BIS SPAN and RV for all the considered stars. The correlation between SN FWHM and SN mean RV is stronger than the correlation between FWHM and RV for three of the four stars. Also for all these stars we corrected the originally estimated RV's from spurious variations in RV's caused by stellar activity, using Eq. 5. Fig. 19–23 show the resulting corrected RV's. While the Normal and SN residuals, once corrected for stellar activity, are comparable for the stars HD192310 and HD10700, the results of the analyses on the star HD215152 (whose CCF's have lower SNR respect to the previous two analyzed stars) suggest that the residuals for the Normal are $0.054\,\mathrm{m\,s^{-1}}$ higher than the residuals retrieved with the SN analysis. Finally, the results of the analysis on Corot 7, whose CCF's have lowest SNR, show that once corrected from stellar activity the residuals from the Normal fit are $0.336\,\mathrm{m\,s^{-1}}$ higher than the residuals retrieved with the SN analysis.

References

G. Anglada-Escudé and R. P. Butler. The harps-terra project. i. description of the algorithms, performance, and new measurements on a few remarkable stars observed by harps. *The Astrophysical Journal Supplement Series*, 200(2):15, 2012.

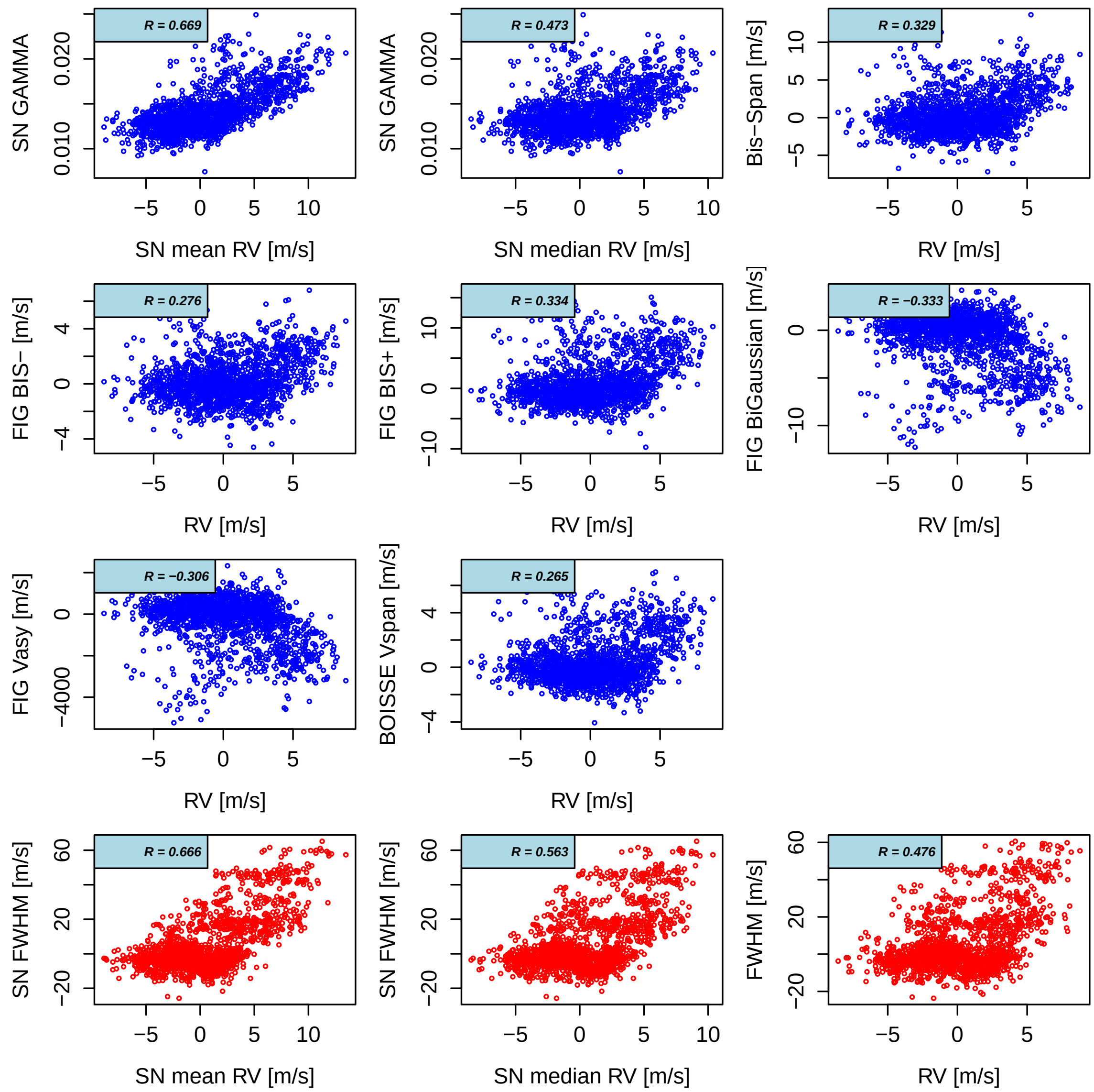

Figure 18: Correlation between the asymmetry parameters and the RV's for HD192310. The last three plots show the correlation between the FWHM and the RV's for HD192310 using respectively the SN and the Normal fits. The p–values associated with each R is statistically different from 0.

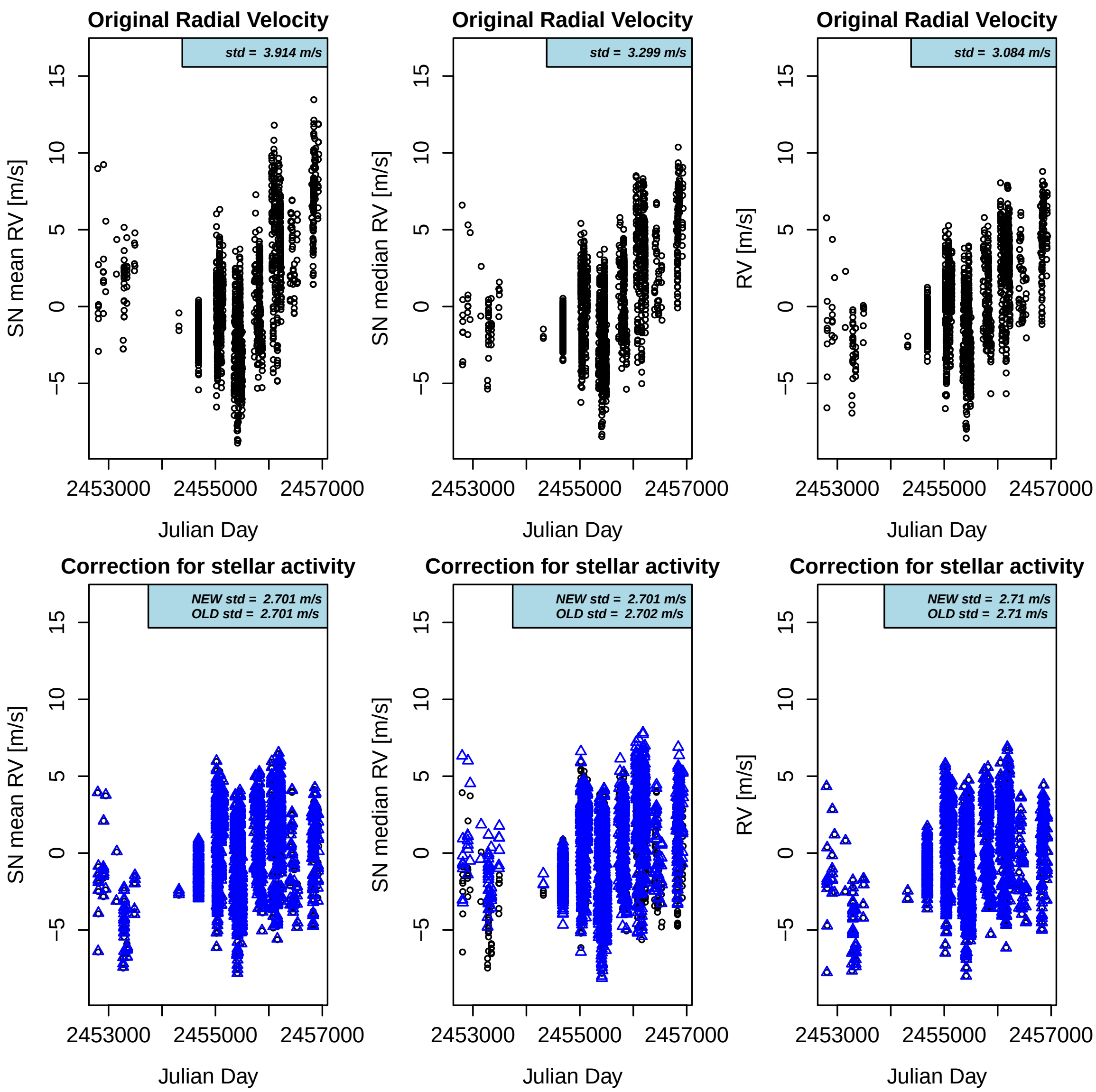

Figure 19: Set of RV's for HD192310 using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. 5. Once corrected for stellar activity, the residuals in the Normal and SN analyses are comparable.

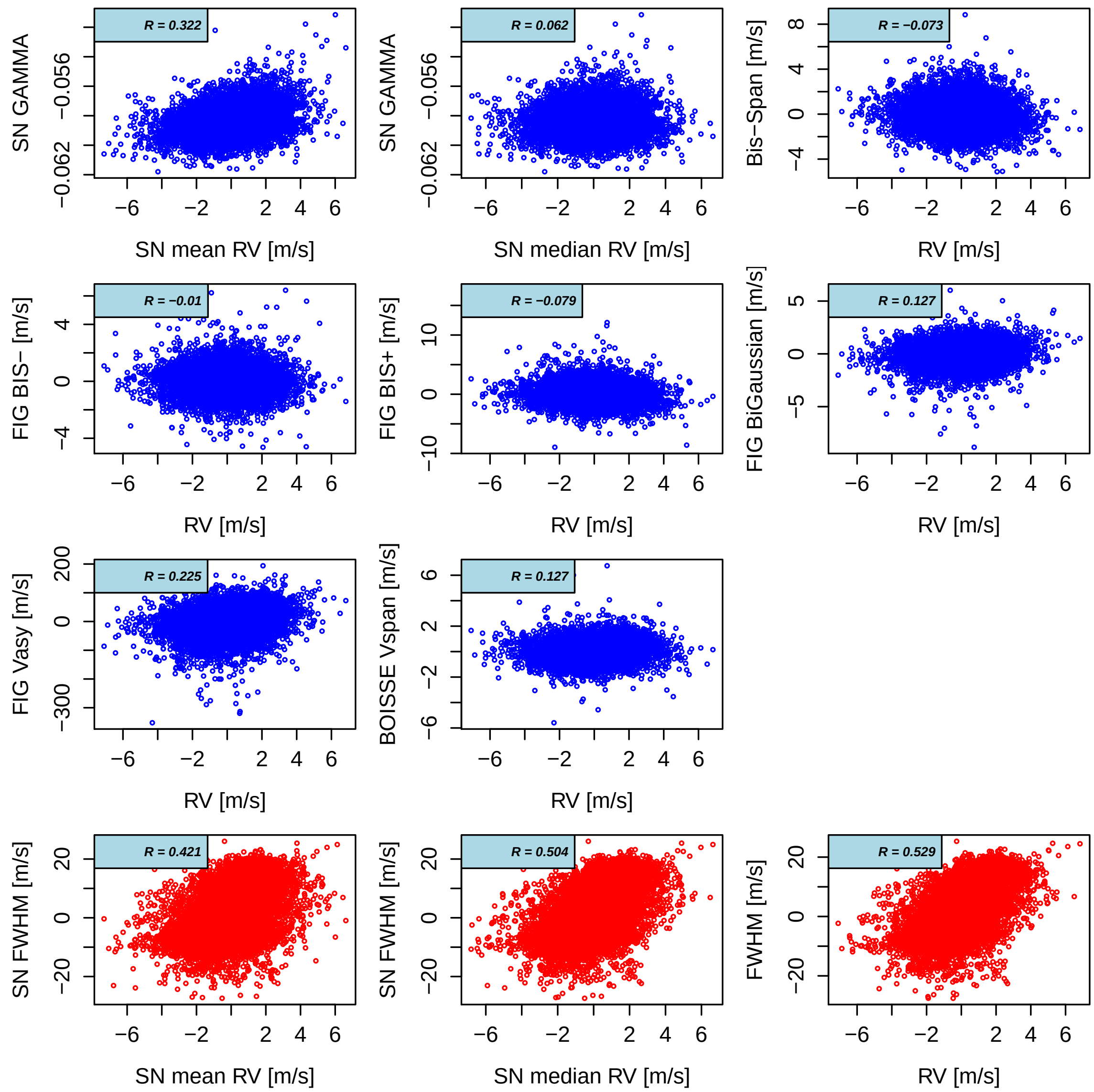

Figure 20: Correlation between the asymmetry parameters and the RV's for HD10700. The last three plots show the correlation between the FWHM and the RV's for HD10700 using respectively the SN and the Normal fits. The p–values associated with each R is statistically different from 0.

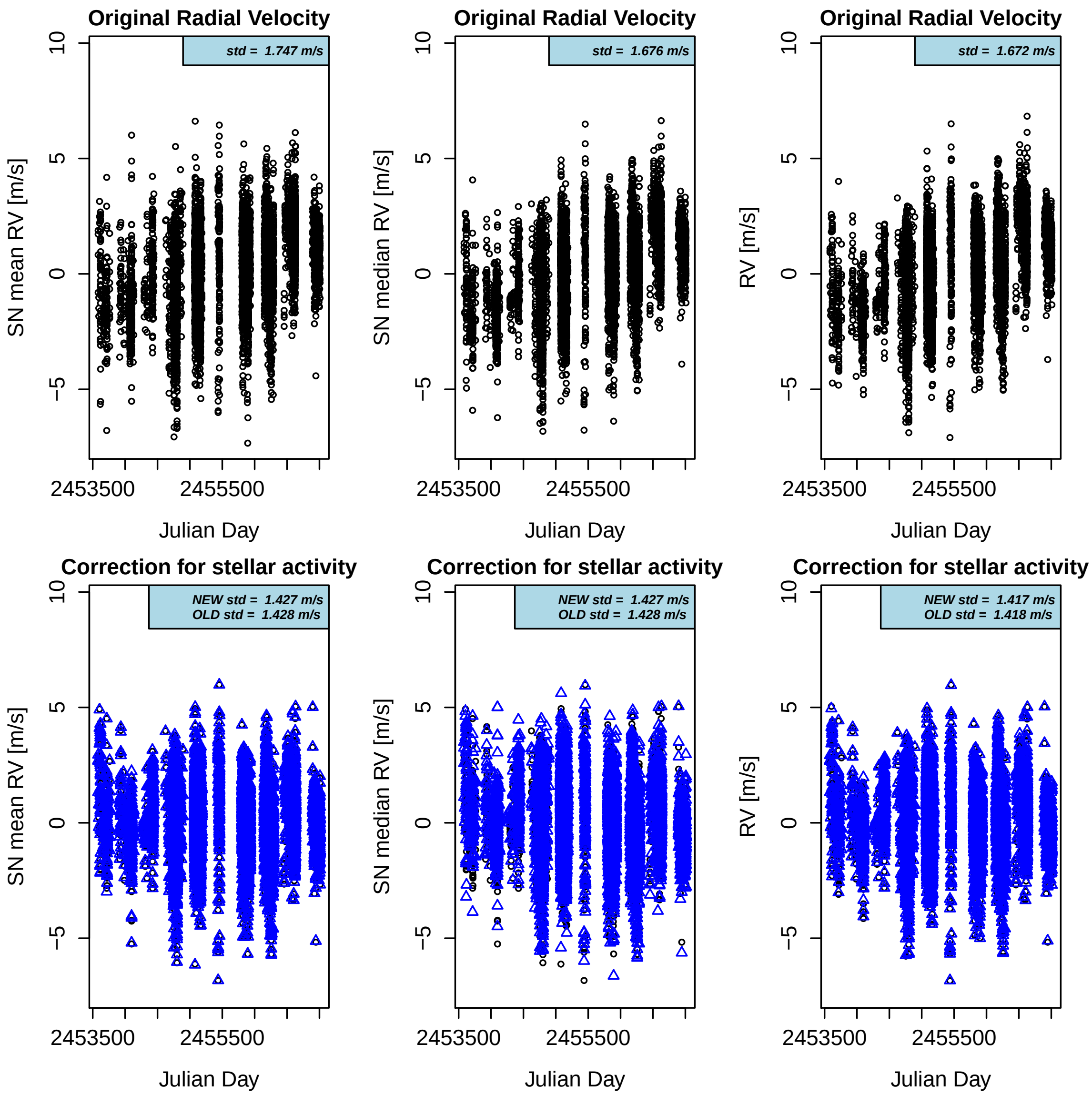

Figure 21: Set of RV's for HD10700 using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. 5. Once corrected for stellar activity, the residuals in the Normal and SN analyses are comparable.

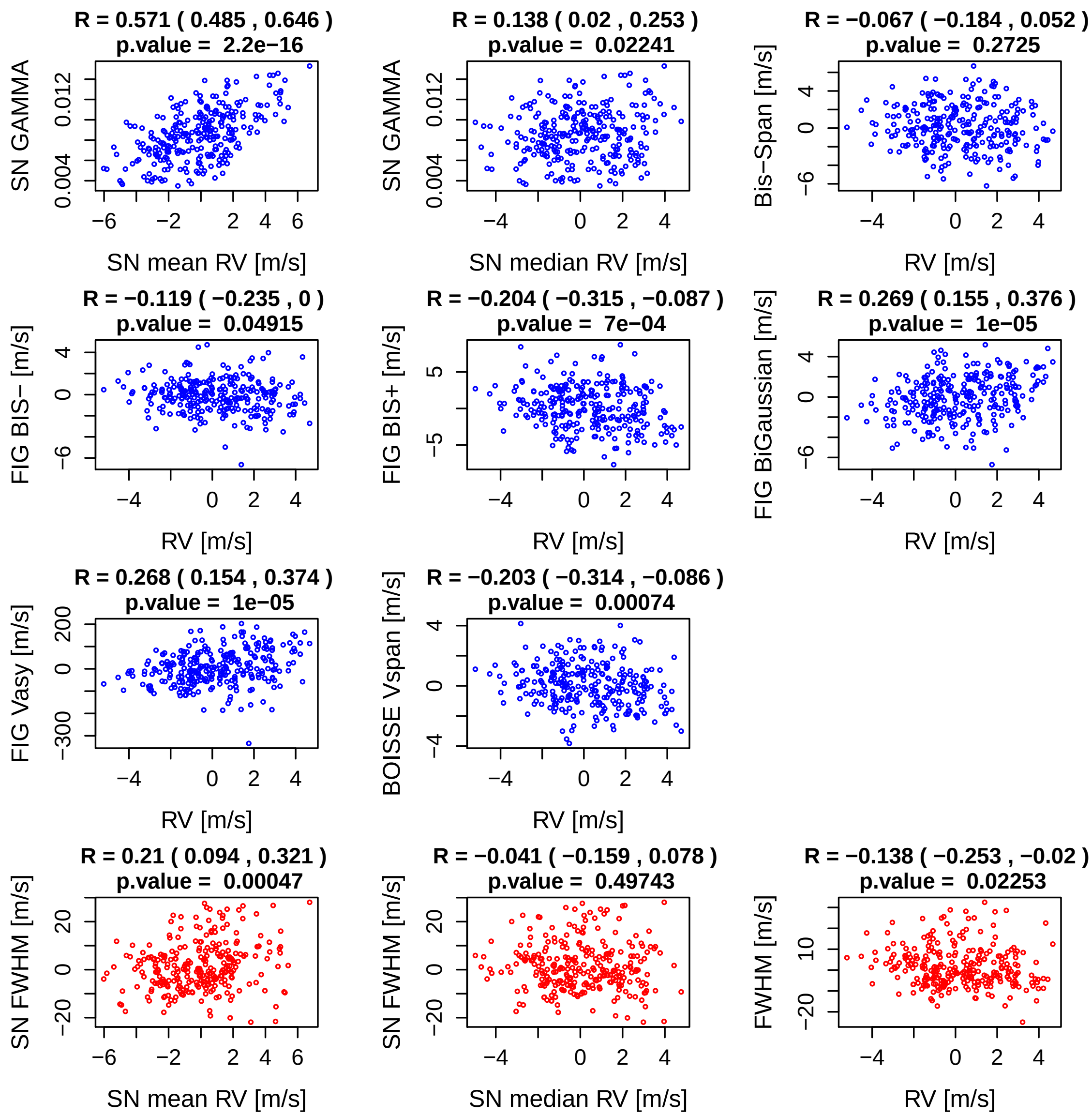

Figure 22: Correlation between the asymmetry parameters and the RV's for HD215152. The last three plots show the correlation between the FWHM and the RV's for HD215152 using respectively the SN and the Normal fits. Concerning the asymmetry of the CCF, note that the p–values associated with R are strongly different fro 0 for those parameters retrieved by using the SN.

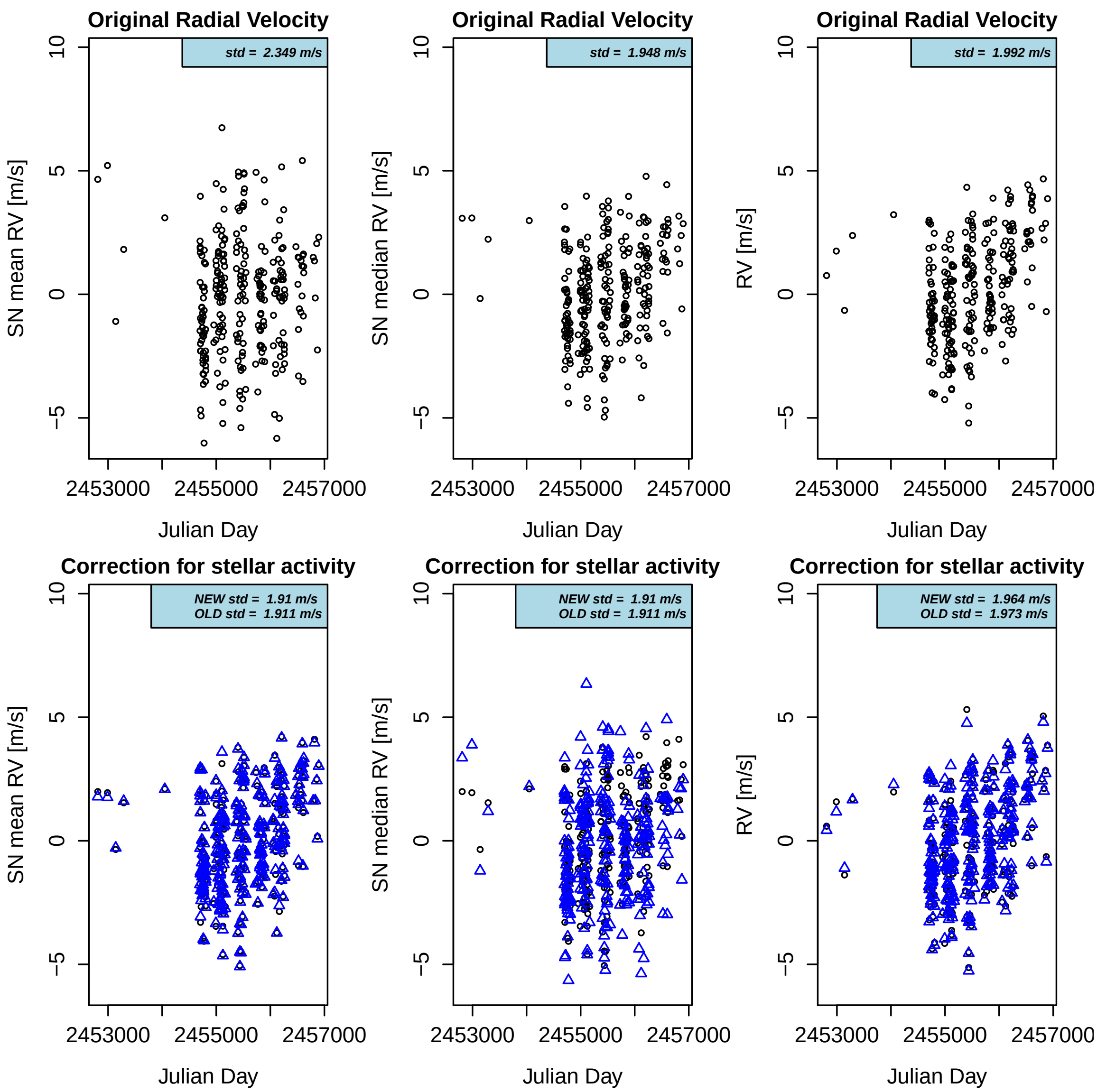

Figure 23: Set of RV's for HD215152 using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. 5. Once corrected for stellar activity, the residuals for the Normal are 0.054 $\mathrm{m\,s^{-1}}$higher than the residuals retrieved with the SN analysis.

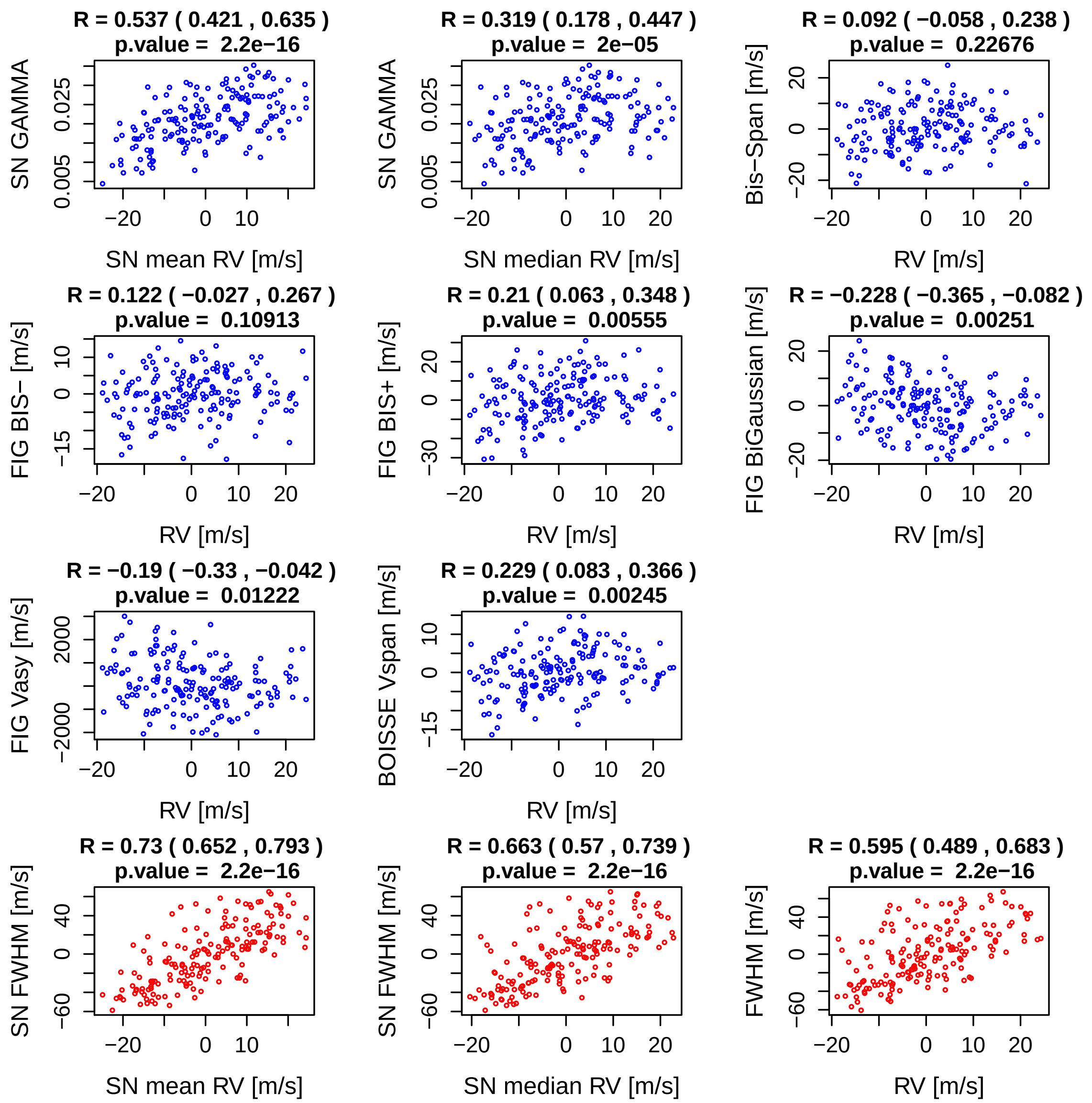

Figure 24: Correlation between the asymmetry parameters and the RV's for Corot 7. The last three plots show the correlation between the FWHM and the RV's for Corot 7 using respectively the SN and the Normal fits. Concerning the asymmetry of the CCF, note that the p–values associated with *R* are strongly different fro 0 for those parameters retrieved by using the SN.

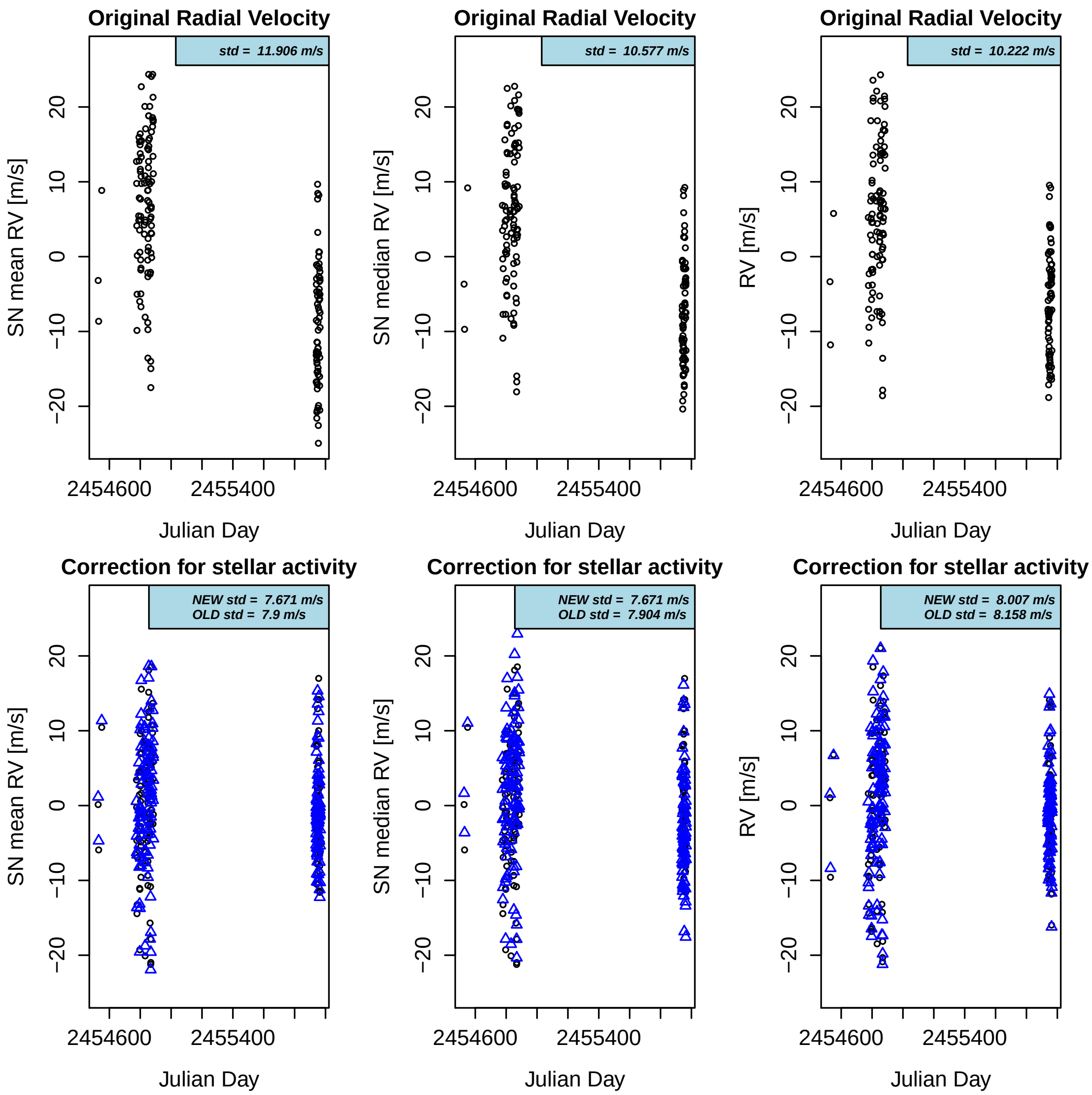

Figure 25: Set of RV's for Corot 7 using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. 5. Once corrected for stellar activity, the residuals for the Normal are 0.336 $\mathrm{m\,s^{-1}}$higher than the residuals retrieved with the SN analysis.

Star	# CCFs	R(SN γ, Bis-Span)	slope(SN γ, Bis-Span)	R(SN γ, SN mean RV)	R(Bis-Span, RV)	R(FIG BiGaussian, RV)	R(SN FWHM, SN mean RV)	R(FWHM, RV)
HD192310	1577	0.888	0.786	0.669(0.64; 0.695)	0.329(0.285; 0.373)	−0.333(−0.376; −0.289)	0.666(0.637; 0.692)	0.476(0.43670.514)
HD10700	7928	0.78	0.604	0.322(0.302; 0.342)	−0.073(−0.095; −0.0051)	0.127(0.105; 0.148)	0.421(0.403; 0.439)	0.529(0.513; 0.545)
HD215152	273	0.763	0.794	0.571(0.485; 0.646)	−0.067(−0.184; 0.052)	0.269(0.155; 0.376)	0.210(0.094; 0.321)	−0.138(−0.253; −0.020)
Corot 7	173	0.814	0.607	0.561(0.450; 0.656)	0.092(−0.058; 0.238)	−0.335(−0.228; −0.082)	−0.709(0.626; 0.776)	0.595(0.489; 0.683)

Table 6: Subset of notable correlations between the asymmetry parameter (and the FWHM) and the RVs for four stars: HD192310, HD10700, HD215152 and Corot 7. The complete results of the analyses of the correlations for the four stars are presented in Fig. 18–24.

R. B. Arellano and A. Azzalini. The centred parametrization for the multivariate skew-normal distribution. 2010.

A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178, 1985.

A. Azzalini and A. Capitanio. The skew-normal and related families. institute of mathematical statistics monographs, 2014.

A. Baranne, D. Queloz, M. Mayor, G. Adrianzyk, G. Knispel, D. Kohler, D. Lacroix, J.-P. Meunier, G. Rimbaud, and A. Vin. Elodie: A spectrograph for accurate radial velocity measurements. *Astronomy and Astrophysics Supplement Series*, 119(2):373–390, 1996.

D. A. Belsley. *Conditioning diagnostics: Collinearity and weak data in regression*. Number 519.536 B452. Wiley New York, 1991.

I. Boisse, C. Moutou, A. Vidal-Madjar, F. Bouchy, F. Pont, G. Hébrard, X. Bonfils, B. Croll, X. Delfosse, M. Desort, et al. Stellar activity of planetary host star hd 189 733. *Astronomy & Astrophysics*, 495(3): 959–966, 2009.

I. Boisse, F. Bouchy, G. Hébrard, X. Bonfils, N. Santos, and S. Vauclair. Disentangling between stellar activity and planetary signals. *Astronomy & Astrophysics*, 528:A4, 2011.

A. Claret and S. Bloemen. Gravity and limb-darkening coefficients for the Kepler, CoRoT, Spitzer, uvby, UBVRIJHK, and Sloan photometric systems. A&A, 529:A75, May 2011. doi: 10.1051/0004-6361/201116451.

R. Cosentino, C. Lovis, F. Pepe, A. C. Cameron, D. W. Latham, E. Molinari, S. Udry, N. Bezawada, M. Black, A. Born, et al. Harps-n: the new planet hunter at tng. In *Proc. SPIE*, volume 8446, page 84461V, 2012.

A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*, volume 1. Cambridge university press, 1997.

M. Desort, A.-M. Lagrange, F. Galland, S. Udry, and M. Mayor. Search for exoplanets with the radial-velocity technique: quantitative diagnostics of stellar activity. *Astronomy & Astrophysics*, 473(3):983–993, 2007.

X. Dumusque. Radial velocity fitting challenge-i. simulating the data set including realistic stellar radial-velocity signals. *Astronomy & Astrophysics*, 593:A5, 2016.

X. Dumusque, S. Udry, C. Lovis, N. C. Santos, and M. Monteiro. Planetary detection limits taking into account stellar noise-i. observational strategies to reduce stellar oscillation and granulation effects. *Astronomy & Astrophysics*, 525:A140, 2011.

X. Dumusque, F. Pepe, C. Lovis, D. Ségransan, J. Sahlmann, W. Benz, F. Bouchy, M. Mayor, D. Queloz, N. Santos, et al. An earth-mass planet orbiting [agr] centauri b. *Nature*, 491(7423):207–211, 2012.

X. Dumusque, I. Boisse, and N. Santos. Soap 2.0: A tool to estimate the photometric and radial velocity variations induced by stellar spots and plages. *The Astrophysical Journal*, 796(2):132, 2014.

X. Dumusque, F. Borsa, M. Damasso, R. F. Diaz, P. Gregory, N. Hara, A. Hatzes, V. Rajpaul, M. Tuomi, S. Aigrain, et al. Radial-velocity fitting challenge-ii. first results of the analysis of the data set. *Astronomy & Astrophysics*, 598:A133, 2017.

B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

M. Efroymson. Multiple regression analysis. *Mathematical methods for digital computers*, pages 191–203, 1960.

F. Feng, M. Tuomi, and H. R. Jones. Evidence for at least three planet candidates orbiting hd 20794. *Astronomy & Astrophysics*, 605:A103, 2017.

P. Figueira, N. Santos, F. Pepe, C. Lovis, and N. Nardetto. Line-profile variations in radial-velocity measurements-two alternative indicators for planetary searches. *Astronomy & Astrophysics*, 557:A93, 2013.

D. A. Fischer, G. Anglada-Escude, P. Arriagada, R. V. Baluev, J. L. Bean, F. Bouchy, L. A. Buchhave, T. Carroll, A. Chakraborty, J. R. Crepp, et al. State of the field: extreme precision radial velocities. *Publications of the Astronomical Society of the Pacific*, 128(964):066001, 2016.

A. P. Hatzes. Starspots and exoplanets. *Astronomische Nachrichten*, 323(3-4):392–394, 2002.

R. Haywood, A. Collier Cameron, D. Queloz, S. Barros, M. Deleuil, R. Fares, M. Gillon, A. Lanza, C. Lovis, C. Moutou, et al. Planets and stellar activity: hide and seek in the corot-7 system? *Monthly notices of the royal astronomical society*, 443(3):2517–2531, 2014.

R. R. Hocking. A biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics*, 32(1):1–49, 1976.

M. Kurster, M. Endl, F. Rouesnel, S. Els, A. Kaufer, S. Brillant, A. Hatzes, S. Saar, and W. Cochran. The low-level radial velocity variability in barnard's star (= gj 699). secular acceleration, indications for convective redshift, and planet mass limits. *ASTRONOMY AND ASTROPHYSICS-BERLIN-*, 403(3): 1077–1088, 2003.

A.-M. Lagrange, M. Desort, and N. Meunier. Using the sun to estimate earth-like planets detection capabilities-i. impact of cold spots. *Astronomy & Astrophysics*, 512:A38, 2010.

S. N. Lahiri. *Resampling methods for dependent data*. Springer Science & Business Media, 2013.

L. Lindegren and D. Dravins. The fundamental definition of ?radial velocity? *Astronomy & Astrophysics*, 401(3):1185–1201, 2003.

M. Mayor, F. Pepe, D. Queloz, F. Bouchy, G. Rupprecht, G. Lo Curto, G. Avila, W. Benz, J.-L. Bertaux, X. Bonfils, T. Dall, H. Dekker, B. Delabre, W. Eckert, M. Fleury, A. Gilliotte, D. Gojak, J. C. Guzman, D. Kohler, J.-L. Lizon, A. Longinotti, C. Lovis, D. Megevand, L. Pasquini, J. Reyes, J.-P. Sivan, D. Sosnowska, R. Soto, S. Udry, A. van Kesteren, L. Weber, and U. Weilenmann. Setting New Standards with HARPS. *The Messenger*, 114:20–24, Dec. 2003.

N. Meunier, M. Desort, and A.-M. Lagrange. Using the sun to estimate earth-like planets detection capabilities-ii. impact of plages. *Astronomy & Astrophysics*, 512:A39, 2010.

M. Oshagh, I. Boisse, G. Boué, M. Montalto, N. C. Santos, X. Bonfils, and N. Haghighipour. SOAP-T: a tool to study the light curve and radial velocity of a system with a transiting planet and a rotating spotted star. A&A, 549:A35, Jan. 2013. doi: 10.1051/0004-6361/201220173.

F. Pepe, M. Mayor, F. Galland, D. Naef, D. Queloz, N. Santos, S. Udry, and M. Burnet. The coralie survey for southern extra-solar planets vii-two short-period saturnian companions to hd 108147 and hd 168746. *Astronomy & Astrophysics*, 388(2):632–638, 2002.

F. Pepe, P. Molaro, S. Cristiani, R. Rebolo, N. Santos, H. Dekker, D. Mégevand, F. Zerbi, A. Cabral, P. Di Marcantonio, et al. Espresso: The next european exoplanet hunter. *Astronomische Nachrichten*, 335(1):8–20, 2014.

D. Queloz, G. Henry, J. Sivan, S. Baliunas, J. Beuzit, R. Donahue, M. Mayor, D. Naef, C. Perrier, and S. Udry. No planet for hd 166435. *Astronomy & Astrophysics*, 379(1):279–287, 2001.

D. Queloz, F. Bouchy, C. Moutou, A. Hatzes, G. Hébrard, R. Alonso, M. Auvergne, A. Baglin, M. Barbieri, P. Barge, et al. The corot-7 planetary system: two orbiting super-earths. *Astronomy & Astrophysics*, 506 (1):303–319, 2009.

V. Rajpaul, S. Aigrain, M. A. Osborne, S. Reece, and S. Roberts. A gaussian process framework for modelling stellar activity signals in radial velocity data. *Monthly Notices of the Royal Astronomical Society*, 452(3):2269–2291, 2015.

P. Robertson, S. Mahadevan, M. Endl, and A. Roy. Stellar activity masquerading as planets in the habitable zone of the m dwarf gliese 581. *Science*, page 1253253, 2014.

S. H. Saar and R. A. Donahue. Activity-related radial velocity variation in cool stars. *The Astrophysical Journal*, 485(1):319, 1997.

A. Thompson, C. Watson, E. de Mooij, and D. Jess. The changing face of α centauri b: probing plage and stellar activity in k dwarfs. *Monthly Notices of the Royal Astronomical Society: Letters*, 468(1):L16–L20, 2017.

D. Wilks. Resampling hypothesis tests for autocorrelated fields. *Journal of Climate*, 10(1):65–82, 1997.